University of Zurich[UZH]

The Perception of Urban Change: A Computational Text Analysis of Newspaper Articles in Zurich North from 1991-2024

GEO 511 Master's Thesis

Author
Lynn Bolliger
18-717-322

Supervised by
Prof. Dr. Ross Purves
Prof. Dr. David Kaufmann (david.kaufmann@ethz.ch)

Faculty representative
Prof. Dr. Ross Purves

23.09.2024
Department of Geography, University of Zurich

Abstract

Research on gentrification has shown that it often leads to the displacement of vulnerable groups of people. The media matters for shaping how the public perceives such social phenomena and thus is an important source to study. This master's thesis investigates how newspapers reported on urban development and gentrification in Zurich North, a rapidly growing region in the city of Zurich, from 1991 to 2024. The study applies Natural Language Processing (NLP) techniques on a large corpus of newspaper articles to detect discourses and changes in reporting over time. The study addresses two goals: first, how useful NLP tools are to investigate media discourse on urban issues based on the example of gentrification. Secondly, how newspapers have reported on gentrification and urban transformation processes in Zurich North within the investigated timeframe. Thereby, the thesis applies frequency and co-occurrence analysis, collocation as well as topic modelling to the corpus to analyse the coverage of key topics over time. To do this neutrally, positively and negatively connotated keywords on the topic were defined and investigated, namely new build, renovation, uplifting, gentrification and displacement. Furthermore, the results of the frequency and co-occurrence analysis are compared to the actual building activity at each of the different time periods to find out whether it correlates with media coverage. The thesis shows that the media regularly reports on construction and renovation activity in Zurich North and that they do so in a neutral to positive manner while only very rarely pointing out the negative effects these projects can have. Thus, the media only contributes little to raise awareness of the negative issues of urban development in Zurich North such as displacement. It was also found that the media coverage matches partially with the construction activity at different time periods. Additionally, the study highlights the importance of context, close reading and careful pre-processing in NLP analysis. The findings of this thesis demonstrate that NLP methods have great potential to study urban issues while also reflecting the limitations and challenges that come with it.

Keywords: Gentrification, Computational Text Analysis, Natural Language Processing

Acknowledgement

First and foremost, I would like to thank Prof. Dr. Ross Purves for all his invaluable support and generous time he gave me throughout the entire process of my master's thesis. He was always there and provided just the right amount of support that I needed to make this master thesis happen.

A special thanks also goes to my co-supervisor Prof. Dr. David Kaufmann for his support, the great thematical inputs he gave me and the interest he showed for my work.

Furthermore, I would like to thank Asli Yaman for supporting me in coding and especially the topic modelling, Magda Breyer and Livia Hess for proofreading and giving me valuable feedback and all my friends who accompanied me through the highs and lows of my studies.

Last but not least, I want to say a big thank you to my family for always being there for me and believing in me. Especially, I want to thank my father, Hansjörg Bolliger who helped me moving forward when I was stuck and proofreading my thesis.

Table of Contents

List of Figures

List of Tables

Abbreviations

BOW Bag-of-words

FSO Federal Statistics Office

GIS Geographic Information System

GWR Swiss Federal Housing and Building Register

MFO Maschinen Fabrik Oerlikon (Machine Factory Oerlikon)

NLP Natural Language Processing

POI Point of Interest

1. Introduction

Today, more than half of the world's population live in urban areas and by 2050 this number is expected to increase to almost 70% (The World Bank, 2023). In many cities around the globe this rapid urbanisation has led to a housing shortage and increasing rent prices. Additionally, in the Global North the housing shortage is also driven by an ageing population and a change of lifestyle: as the household size decreases and more single-person households exist, the demand for housing rises. To lessen the housing shortage, more housing must be provided (Teller, 2021).

The Sustainable Development Goal 11.1 asks to "ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums" (SDG 11.1). Thereby, in many countries city compaction and densification is seen as the "pathway to more sustainable development" (Pont *et al.*, 2021, p. 379) as it counters problems such as climate change, land consumption, urban sprawl and loss of biodiversity (Debrunner and Hartmann, 2020; Pont *et al.*, 2021; Teller, 2021). While urban densification can have many positive effects it can also lead to negative effects on the social life such as displacement and loss of social networks (Teller, 2021). Since many cities lack free inner-city plots, densification is often implemented through redevelopment of existing housing stocks such as total replacement construction, renovation or compaction of existing buildings (Debrunner and Hartmann, 2020). The modernisation of old building stock through replacement constructions and renovations, however, often leads to rising rents from which particularly low-income, old-aged, migrant persons and families suffer as they can no-longer afford the rents where they lived before and are pushed to the periphery or places with fewer services (Debrunner and Hartmann, 2020; Teller, 2021; Kaufmann *et al.*, 2023). Densification can further be linked to increasing land competition, land value extraction, real estate speculation, housing inequalities and new-build gentrification (Cavicchia, 2023). Some studies have shown that densification areas are marketed at and attractive to urban elites who seek "low carbon" urban living and who can afford to live in redeveloped central neighbourhoods, (Cavicchia, 2023, p. 1475). Thereby, well-educated members from the "new" middle-class take over former walkable working-class districts (Rice *et al.*, 2020). This process is called gentrification and has been studied extensively since the 1960s (Slater, 2011).

Gentrification studies are often done with traditional qualitative methods such as interviews or quantitative analyses of census data over time to identify changes in the ethnic composition, income and educational attainment of residents (Loukaitou-Sideris, Gonzalez and Ong, 2019). Qualitative interviews allow to get an in-depth view about past and current developments as well as individuals' feeling towards it. However, they are often quite resource and time intensive. Quantitative census-based methods on the other hand take up less resources but rely on census data updates. Further, they allow an insight into where and when gentrification take place but not on how people feel about it.

While urban densification is important to create enough and more sustainable housing, it is of upmost importance to be aware of the negative effects it can have especially on housing affordability and displacement. Only if there is awareness around the issue, measures can be taken early to lessen the negative social effects. Newspapers and media more generally are an important seismograph for public opinion and take large parts in shaping the public discourse. Regarding urban development, they can help to raise awareness of what is going on in the city and feel the temperature of the population on how they feel about the ongoing processes as well as potential undesirable developments. It is therefore important to know if, where and how newspapers report on the issue of gentrification and displacement. Analysing textual data, such as newspaper articles, in a discourse analysis is a popular method used in many different fields of research such as human geography, linguistics or political science and is traditionally conducted in a qualitative manner (Johnson and McLean, 2020). These methods allow thorough analyses of text, provide valuable insights and have also been applied to study

how newspapers report on gentrification (e.g. Brown-Saracino and Rumpf, 2011; Tolfo and Doucet, 2021). However, reading many articles can be very time consuming especially when studying long time periods or large geographical areas. In recent years, vast numbers of text have become available digitally, opening the field to quantitative text analysis. These methods have come to be known under the term Natural Language Processing (NLP) tools and are aimed at analysing text to solve a variety of challenges: from statistical analyses of word co-occurrence to translating text (Purves, Koblet and Adams, 2022b). NLP allows to analyse large number of texts within a relatively short time and thus bridges the gap between the valuable insights gained from text analyses and the feasibility to study large geographical areas and long time periods.

1.1 Objectives and Research Questions

The aim of this thesis is twofold: on one hand, I want to explore in what ways NLP tools are useful to analyse the newspaper discourse around gentrification processes. Thereby, I intend to apply three different NLP tools, namely, frequency and co-occurrence, collocation and topic modelling to understand how they can help to better understand the processes as well as identify their limitations. On the other hand, I want to explore how newspapers reported on gentrification and its effects in the northern Districts of Zurich, Switzerland since 1991. The focus will be on where newspapers discuss issues of gentrification and how it relates to the actual building activity in the region at that time. To do so, I intend to compare the locations often mentioned in the newspapers with the Swiss Federal Housing and Building Register (GWR). Furthermore, I will explore the context of certain keywords related to the process of gentrification and places of interest to investigate how they are used and if they have changed over time.

This thesis seeks to close the research gap of lacking research on gentrification in Zurich North, a fast growing, dynamic region in the city of Zurich. Additionally, the thesis intends to bring together discourse analysis of gentrification in newspaper articles with modern NLP tools. A lot of research exists already on both topics separately. However, as far as I could see no study so far uses newspapers to find discourse on gentrification with NLP methods. The most similar paper found was by Jiang et al. (2023) who leveraged newspapers to understand urban shrinkage in Detroit. The advantage of using newspaper articles over social media data is that it reaches far beyond the last 10 – 20 years, for which social media data exist. As many urban phenomena such as urban growth or gentrification unfold over a long period of time, social media data can only provide limited insights over a short timeframe (Jiang *et al.*, 2023). Additionally, while many social media platforms are restricting the access to their data for researchers (Davidson, Hinds and Racek, 2023), more and more newspaper are getting digitalised, further expanding the time frame for analyses (Jiang *et al.*, 2023). Specifically, this thesis seeks to answer two specific research questions:

> **RQ 1:** In what ways can NLP tools be used to analyse discourse around gentrification and displacement over time in a newspaper corpus?
>
> **RQ 2:** How do newspaper talk about gentrification and urban development in the northern districts of Zurich?

1.2 Structure

Chapter 2 “Background” examines the background of gentrification such as the definition, causes and effects as well as a short overview about gentrification studies in Zurich. Further, Chapter 2 looks at Natural Language Processing in general and describes different tools that will be used in this thesis. In a last part, different examples of academic studies that used NLP tools to study urban spaces are presented. Chapter 3 “Methods” describes the data, the research area, the pre-processing approaches and the methods used in this thesis in detail. In Chapter 4 “Results”, all the results that were generated with the three methods are discussed separately. Based on these results, the research questions are answered in Chapter 5 “Discussion”. Furthermore, it will summarise what has been achieved in this thesis, the limitations and weaknesses of the study as well as the implications the results have on urban studies. Finally, Chapter 6 “Conclusion” completes the thesis by reviewing the most important findings and implications of the thesis.

2. Background

2.1 Gentrification

Gentrification is a well-studied subject and a process that has been observed in cities around the globe, both in the Global North as well as the Global South (Slater, 2012). The word “gentrification” has been coined by sociologist Ruth Glass in the 1960s when she was describing the “invasion” of the middle-class in working class quarters in London leading to the displacement of the working class and the changing of the social character of the district (Slater, 2012). Since then, the term has been widely used to describe the process of renewal and rebuilding of inner-city urban areas accompanying an influx of young, educated, high-income residents in formerly low-income neighbourhoods (Zenebe, Brown-Robertson and Mayo, 2018). However, even though the term is widely used, there is no precise definition of it. Until the 1980s gentrification was mostly used according to Ruth Glass’ definition (Slater, 2012; Naismith and Murphy, 2023). However, since then the definition has been expanded to include other processes that lead to the direct and indirect displacement of people and new terms such as rural gentrification (Phillips *et al.*, 2021) or new-build gentrification on vacant or former industrially used areas have emerged (Davidson and Lees, 2005, 2010; Rérat et al., 2009; Slater, 2012).

2.1.1 Causes of gentrification

Gentrification is a complex and dynamic process that does not play out the same everywhere and which cause-and effects relations can be unclear. In the past there have been two main strands of possible explanations: the production-side theory versus the consumption-side theory. One of the most important theories of the production-side argument is the rent-gap theory developed by Neil Smith in the 1980s (Slater, 2012; Pegler, Li and Pojani, 2020). The rent gap describes the gap between the actual rent paid for a piece of land and the potential rent if the land had a “higher and better use” (Smith, 1987:462). To close the rent gap, landlords can renovate, redevelop or transform the existing infrastructure and usage to a “higher or better use”. The uplifting of this land paves the way for the middle- and upper class to move into these areas and transforms them further to accommodate their needs. The renovation of houses as well as the influx of new residents increases the rents which can lead to the displacement of the former residents. The explanatory theory from a consumption perspective sees gentrification as the result of a demographic change and with it a change in lifestyle preferences, tastes and desires of the middle class (Pegler, Li and Pojani, 2020). Thereby, the new middle class prefers the inner-city location over the suburbia as it provides a wide offer of good jobs, recreational and cultural amenities as well as the opportunity to meet a wide variety of people, which

allows them to express their personal style (Ley, 1986). Furthermore, the scholars of the consumption perspective see a change in the economic structure of cities from an industrial to a service-based economy. This shift leads to a change from blue collar workers to a "rapidly growing downtown workforce […] of private and public corporate employees, professionals, university and hospital staff, and those engaged in the arts and media" (Ley, 1986, p. 525) which make up the new middle class with a new, distinctive lifestyle. The lifestyle of this new middle class also manifests in the physical environment with the emergence of craft breweries, coffee shops and yoga studios. While in the 80s these two explanatory strands were considered to be opposite of each other, it is now acknowledged that both of them can contribute to understand the causes and processes of gentrification and are thus supplementary (Slater, 2011).

Another influential theory regarding gentrification is the "wave" theory developed by Jason Hackworth in the 2000s (Pegler, Li and Pojani, 2020). Hackworth observed that gentrification is occurring in waves whereby the timeline mostly follows the development in the United States and more specifically in New York but still applies to many countries in the global North (Hackworth and Smith, 2001). According to Hackworth & Smith (2001), the first wave of gentrification happened in the 1960s and early 70s and can be described as sporadic, highly localised and state-led: governments around the global North reinvested in inner-city neighbourhoods to stop urban decline as the private sector economy was in a downturn phase and disinvested in these areas. The second wave of gentrification set in at the end of the 1970s and lasted until the end of the 1980s and is defined by Hackworth and Smith as "expansion and resistance". In this wave it was mostly the private market that profited from a "laisser-faire" role of the state that redeveloped many new neighbourhoods also in smaller cities that were so far untouched by the process. In this wave, the presence of alternative arts scene was often observed as a correlator to gentrification. Furthermore, it was at this time that criticism rose against gentrification as it led to the eviction of already vulnerable residents and homelessness. The second wave of gentrification was halted by the economic recession of the early 1990s but soon, in the mid-1990s the third wave set in. This wave is characterised, according to Hackworth and Smith (2001) by an even larger expansion to neighbourhoods outside of the inner-city areas, the emergence of large, global developers, less resistance and a more active role by the state (Hackworth and Smith, 2001). The idea of the gentrification process as waves was quite influential and shaped the discussions until today. Aalbers (2019) picked up on the three waves of gentrification and argues that after the global financial crisis in 2008, a fifth wave gentrification process set in that is characterised by "corporate landlords, highly leveraged housing, platform capitalism (e.g. Airbnb), transnational wealth elites using cities as a 'safe deposit box', and a further 'naturalisation' of state-sponsored gentrification" (Aalbers, 2019:1).

2.1.2 Effects

While it is important to understand the causes of gentrification it might be even more important to analyse the effects it can have. Some believe that gentrification is mostly a good thing as it revitalises formerly rundown neighbourhoods and leads to an economic uplift with better jobs, higher income, more amenities and a more liveable urban environment (Slater, 2011). However, while some of these positive effects can occur, gentrification also often leads to the displacement of former residents which is much criticised. Because neighbourhoods are redeveloped, houses renovated and more affluent persons are attracted to an area, the rents increase. Especially in low-income neighbourhoods many of the residents that have been living in and shaping the neighbourhood, may no longer be able to afford the new rents and are forced to move out of the inner-city neighbourhoods to the city edge or even outside of it. However, displacement does not always have to be direct through the increase of rent but can also occur indirectly through displacement pressure or exclusionary displacement (Marcuse, 1985). Displacement pressure is given when the social networks become fragmented through moving

out of former inhabitants and new amenities that mainly serve middle-class needs. This can lead to an alienation of residents that have been living in a neighbourhood for a long time but can no longer identify with the new image of it (Atkinson, 2015). Exclusionary displacement on the other hand occurs when a household moves out of its residence which is then gentrified and as a consequence "another similar household is prevented from moving in" (Marcuse, 1985, p. 206). Besides these two types of displacement, Marcuse (1985) also identified nine other types of displacement associated with gentrification processes such as social displacement or economic displacement, that are nicely summarised in Phillips et al. (2021). Being forced to move out can be a traumatic experience as these people not only lose their home but also their entire social network (Elliott-Cooper, Hubbard and Lees, 2020). Furthermore, people who have been forced to move, expressed their feelings of loss, regret, bitterness and resentment in interviews and those that stayed that their place attachment was damaged with the gentrification processing, feeling that the new services and amenities were "not for us" (Atkinson, 2015, p. 385).

2.1.3 Housing shortages, increasing rents and gentrification in Zurich

Discussions on housing shortages and questions of whom the city belongs have occupied the city of Zurich for almost a century (Stahel, 2006). Stahel who studied social housing movements in Zurich for his dissertation (2006), concluded that in Zurich, the housing shortage has become an existential problem for broad sections of the population in regular 10-year cycles. Reasons for the increasing intensity of housing shortages are manifold.

There are two factors that mainly contribute to the intensity of the housing shortage: the number of people seeking to live in a city and demographics. Surprisingly it is not primarily the total number of inhabitants that influences a housing shortage the most, but the population composition: with an increase of 20–29-year-olds the demand for living room increases even if the total population decreases, as they are most likely to establish their own household (Stahel, 2006). This was shown during the 1980s where there was an acute housing shortage even though the population of Zurich has been decreasing since the 1960s as more and more families moved to rural areas, no longer being able to afford flats in the inner city and seeking the advantage of a rural live (Stahel, 2006). Furthermore, the household size and demands impact the availability of housing: over the last decades the household size decreased, having more single person households and an observed increase in living area per person (Stahel, 2006). For example, the average living space in the city of Zurich increased from 29.6m^2 in 1970 to 41.2m^2 in 2022 (City of Zurich, 2023e) and the number of single households increased from under 20% in 1960 to around 45% in 2013 (City of Zurich, 2014). The accumulation of these effects can increase the housing shortage even if new housing is built and the population decreases (Stahel, 2006). However, since the 1990 the population in the City of Zurich is steadily growing, having broken the all-time population record from 1962 in 2022 (Rey, 2022). Today almost 450'00 people live in the City of Zurich (City of Zurich, 2024a).

On the other hand, there is the supply side: Since the 1950s the housing building activities have steadily decreased in the City of Zurich especially from the building cooperatives that provide affordable living room (Stahel, 2006). Reasons for that are that land reserves were becoming scarcer as more were overbuilt and that investors preferred building offices instead of housing as it promised more returns leading to the conversion of existing living room to office buildings (Stahel, 2006). This lack of sufficient living space led to rising rents during economic upswings where a lot of people move into the city for work, which increases the demand for housing. Furthermore, long-term strategies for urban housing policies were lacking for a long time in the City of Zurich, so that measures to counter housing shortages were only implemented in reaction to acute shortages. However, as building new living room takes

time, this would lead to an escalation of the housing shortage until the economy has weakened again (Stahel, 2006).

Gentrification in Zurich has most likely first emerged in the district 8 (Seefeld) in the early 1990s. The former redlight district next to the lake was uplifted and the rents increased by 20% within 5 years (NZZ, 2009). Small shops and restaurants made way to luxury boutiques and restaurants. This is why, "Seefeldisierung" today is used as a synonym for gentrification in Zurich and even Switzerland (Sturzenegger and Lutz, 2017). In the mid-1990s the next inner-city neighbourhoods followed: the district 4 and 5, former industrial and working-class neighbourhoods (Stahel, 2006; Craviolini, Heye and Odermatt, 2008). After the end of the open drug scene in Zurich in the mid-1990s, and the disappearance of drug addicts from the city image, the neighbourhoods were quickly uplifted: young people moved into the districts, trendy bars opened, many houses were renovated, and new construction projects were being planned with the support of the city government. At the same time, former industrial areas were being redeveloped (Zurich West) to make way for office spaces, expensive apartments and trendy restaurants (Stahel, 2006). This led to a steady decrease of the share of the foreign working-class population while the socio-economic status of its population increased (Stahel, 2006; Craviolini, Heye and Odermatt, 2008; NZZ, 2009). Another place that is often mentioned in regard to gentrification in Zurich both in newspapers as well as research is the Langstrasse, one of the nightlife districts in Zurich with many trendy bars, clubs and restaurants that changed from a run-down area to one of the trendiest neighbourhoods in Zurich (Craviolini, Heye and Odermatt, 2008; Bürgi, 2014; Widmer and Kübler, 2014; Sturzenegger and Lutz, 2017).

Discussions around gentrification in Zurich have emerged in the early 2000s and never really stopped since. However, they have recently re-emerged more intensively in the last few years with a steadily increasing population and an escalation on rent increases that led to an acute (affordable) housing shortage. On July 1st, 2024, only 169 apartments were empty with a housing vacancy rate of 0.07% (City of Zurich, 2024k). According to experts, the housing market would be in an equilibrium with vacancy rates between 1% and 1.5% (Thalmann, 2012). This shows that the housing market in Zurich is extremely dry, causing month long searches to find suitable apartments and leaving many to look for apartments outside of the city (Jungen, 2023). At the same time many apartments are added each year through renovations, replacement buildings and new-builds as the building activity has been high in recent years (City of Zurich, 2024j). In a recent study, Kaufmann et al. (2023) analysed the building activity, displacement and acceptance in the Canton of Zurich. They found that in the Canton of Zurich most new apartments are added through replacement buildings which leads to the direct displacement of vulnerable groups of people. They discovered that almost 13'000 people have been directly displaced by renovations or new replacement buildings between 2014 to 2019. Most of them move either to the outskirt neighbourhoods in Zurich North (district 11 and 12) or even outside of the city. Thereby, disproportionately vulnerable groups of people are affected such as persons with low-income, foreigners (30% more likely to be affected) and single-parent households (50% more affected) (Kaufmann *et al.*, 2023).

Gentrification in Zurich has so far mostly been discussed around the Langstrasse, Kreis 4, 5 and 8 (Craviolini, Heye and Odermatt, 2008; Rérat *et al.*, 2009). These are inner city locations that fall within the more typical definition of Ruth Glass' definition for gentrification. Less in the focus of research are the districts in the outskirts of the city that are amongst the fastest growing in recent years: Affoltern, Oerlikon, Schwamendingen and Seebach, summarised under the name Zurich North. Most of the housing in Zurich North was built during the 1950s and 1960s for the workers of the industrial sites in Oerlikon (Widmer and Kübler, 2014). Some of the biggest remaining land reserves are also located within Zurich North. The large land reserves as well as the old building substance has contributed to the fast development of Zurich North as for the past twenty years they have been developed and

densified with replacement buildings. A particularity of the districts in Zurich North is the high share of housing owned by housing cooperatives. The aim of housing cooperatives is to provide adequate and affordable housing without making a profit (Wohnbaugenossenschaften Schweiz, 2024). Despite the provision of apartments below private market prices, the building activity of housing cooperatives can still have a negative impact as the rent increases after the renovation and rebuild. In an interview the president of the Swiss Heritage Protection Martin Killias criticises that the social question about the displacement of low-income people is ignored and states that densification has a tragic social component. He expresses his concern that housing cooperative estates are torn down to build a dozen apartments more for double the rent:

> "*Was mich aber traurig macht, ist, dass die soziale Frage, die Verdrängung der einkommensschwachen Menschen, vollständig ignoriert wird. Genossenschaftssiedlungen werden abgebrochen, um einen Block mit einem Dutzend Wohnungen mehr, aber doppelt so teuren Mieten zu bauen. Ich will nicht moralisieren, aber ich kann mich der sozialen Solidarität nicht entziehen. Verdichtung hat eine tragische soziale Komponente*." (Scagliola, 2024)

Considering these large transformations that were and are still happening in Zurich North, it is important to keep an eye on the developments in these neighbourhoods to mitigate the potentially negative impacts they can have on its inhabitants.

2.2 Natural Language Processing

Natural Language Processing (NLP) tools are becoming increasingly popular in research. Since the emergence of the internet and social media the amount of digitally available texts has steadily grown, providing a large treasure of information. Big data is produced in a high volume, velocity and a broad variety and can therefore be used to analyse a large range of topics (Zenebe, Brown-Robertson and Mayo, 2018). Among all existing data, 95% are in unstructured form of flowing text which is challenging to explore for a computer as it lacks an identifiable tabular organisation that is used in traditional data analysis methods (Cai, 2021). NLP is part of artificial intelligence and seeks to close this gap by using computational algorithms and rules to analyse text and its structure by using computational techniques to learn, understand and produce human language content (Hirschberg and Manning, 2015; Cai, 2021). Even though great progress has and is being made in building reliable and efficient systems, there is still a long way to go (Purves, Koblet and Adams, 2022a). Reasons are the ambiguity, variability and context-dependent interpretation of the human language (Hirschberg and Manning, 2015). Aaccording to a conservative estimation at least 32% of the words in the English language are lexically ambiguous (Cai, 2021). Nonetheless, NLP methods are very useful for a variety of tasks as they enable us to statistically analyse written text and reveal patterns, topics and sentiment for large amounts of text a human could not handle (Purves, Koblet and Adams, 2022a). Popular applications of NLP tools are for example, machine translation (translating text from one language to another), spoken dialogue systems (enabling communication between humans and machines e.g. in Apple's Siri), machine reading (summarise information from vast amounts of text), mining social media or sentiment analysis (identifying positive and negative connotations in textual language) (Hirschberg and Manning, 2015).

One of the earliest and most straightforward NLP techniques are the so-called bag-of-words (BOW) models. BOW models ignore the semantic structure, grammar and order of the words and assume that the frequency of words alone is enough to find meaningful information in a corpus (Purves, Koblet and Adams, 2022a). Beside the BOW models other models exist based on machine learning (ML) that consider grammar, sentence structure and word order and allow to identify syntactic and semantic information as well as discover discourse context (Hirschberg and Manning, 2015). They are among the best-performing systems in use (Hirschberg and Manning, 2015). However, they are more complex to

implement and require massive amounts of data, computational resources and energy (Purves, Koblet and Adams, 2022a). Because of this, this thesis mainly uses approaches based on BOW models, which are "simpler, established methods [that] can be quite effective when used appropriately and in practice are often much easier to apply" (Purves, Koblet and Adams, 2022a, p. 56).

An important step in NLP is the pre-processing of the corpus which translates the raw, unstructured text into a form that computers can process. One very important pre-processing step is the so-called tokenisation. Tokenisation is the process of "splitting textual data into smaller meaningful components called tokens" (Sarkar, 2016, p. 108) such as sentences or individual words. Tokenising a corpus facilitates the removal of so-called stopwords and normalisation. Stopwords are words that have "little or no significance" (Sarkar, 2016, p. 120) for the meaning of a text but appear frequently such as "and", "or", "a", "the" or "me". Because these words have generally a high frequency in text documents, they would be predominantly extracted when using BOW models that are based on frequency and statistics. Since they convey very little meaning, these words are often removed in the corpus pre-processing by using stopword lists. Another important pre-processing step is to group all words with the same meaning together so that their frequency is counted correctly by using normalisation, stemming and lemmatisation. Normalisation tries to bring all word forms into a single canonical word form (e.g. "Zurich" and "Zürich") (Purves, Koblet and Adams, 2022a). Lemmatisation and stemming are both normalisation methods that reduce the inflected words to root forms (e.g. the root form of "singing" is "sing") by removing the affixes of words (Sarkar, 2016; Purves, Koblet and Adams, 2022a; Wartena, 2023). While stemming reduces a word to a stem that may not exist as such in a dictionary, the lemma returned by lemmatisation will always be an existing word that can be found in a dictionary (Sarkar, 2016). Sarkar (2016) also recommends to contract expansion such as "can't" to "cannot" and word and spelling correction in the pre-processing workflow. All these steps improve the result output of NLP methods and should thus always be conducted before an analysis.

The simplest kind of BOW analysis is the so-called frequency of a term. The frequency of term can be calculated for single words or n-grams in different documents which can then be compared against each other. Thereby, it assumes that the higher the frequency of a word, the more important it is. However, simple frequency alone does not show how meaningful a term is in a document, but it must be accounted for how common it is in a language in general and in a corpus in particular (Purves, Koblet and Adams, 2022a). One popular method to include the abundance of a word in the corpus is the TF-IDF method (Purves, Koblet and Adams, 2022a). TF-IDF stands for Term Frequency – Inverse Document Frequency and combines two metrics: Term frequency and inverse document frequency. The term frequency *tf* is the number of times a particular word appears in one document. The inverse document frequency is calculated by dividing the total number of articles in the corpus *N* with the number of articles that contain the particular search word, *df*. The TF-IDF score is the result of multiplying the two above measures together. By also considering the document frequency, TF-IDF reduces the bias words with high frequencies in the corpus have over more interesting words with less frequencies (Sarkar, 2016). Thus, words that have a high frequency in one text but low frequencies in the whole corpus are more specific and thus of higher interest than words that are abundant in the corpus.

Formula for TF-IDF score in scikit: $tf \times \frac{\ln(N+1)}{(df+1)} + 1$

Another popular BOW method is the so-called collocation. Collocation is defined as the "combinations of words that habitually co-occur in texts and corpora" (Brezina, 2018:67), therefore it looks at what words are often found in the vicinity of another word. To generate collocations a "node" and a "collocation window" must be defined (Brezina, 2018). The node is the word of interest of which we want to know what other words appear in its vicinity. The collocation window defines in what span from the node, collocations are extracted. The most common types of collocation are bigrams and

trigrams, whereby bigrams extract two adjacent words and trigram three adjacent words (Ruchirawat, 2018). Based on these two inputs a collocation algorithm looks for all possible collocations in a text. However, not all extracted collocations are meaningful or interesting which is why, the observed frequency of the collocates must be evaluated. According to Brezina (2018), there are three options to evaluate the observed frequency: 1. Produce a rank-ordered list based on the observed frequency of the collocate, 2. Compare the observed frequency with the expected frequency based on a random co-occurrence baseline and 3. Compare the observed frequency with a different non-random baseline. The first option is very easy to apply and does not use any statistical calculations. However, because it only looks at the observed frequency of the collocates, the top collocates extracted will typically be generic words that collocate often due to their high frequency in the corpus in general. Therefore, this method only has a limited usefulness (Brezina, 2018). Option number two compares the observed frequency of the collocation against a random baseline. Thereby, it assumes a "shake the box" language model which calculates how often a collocate would co-occur with a node by chance if all the words in the corpus were put into a random order: the expected frequency. This greatly simplifies the complexities and structure of a language but is easy to calculate and can be done without any further inputs. (Brezina, 2018). The last option avoids the "shake the box" model and takes the structure of a language more into account when calculating the expected frequency. However, it also adds a lot of complexity to the evaluation and often needs external input.

The expected frequency based on random co-occurrence can be calculated with the following equation:

$$Expexted\ Frequency = \frac{node\ frequency * collocate\ frequency * window\ size}{Nr\ of\ tokens\ in\ a\ text}$$

The observed frequency of the collocate should then be compared against the expected frequency with association measures which "calculate the strength of association between words based on different aspects [frequency and exclusivity] of the co-occurrence relationship" (Brezina, 2018, p. 67). Depending on the research question, different association measures should be used such as for example the t-test, mutual information or log likelihood. The strength and weaknesses of the different association measures are explained in detail in in the book chapter Semantics and Discourse by Brezina (2018).

Another NLP application is topic modelling. Topic models are designed to extract different themes and concepts from a large number of documents (Sarkar, 2016) and can help to explore corpora in a straightforward way since it does not use any training data or manual annotation (Koblet and Purves, 2020). One very commonly used approach is the Latent Dirichlet Allocation algorithm (LDA) (Koblet and Purves, 2020). LDA is a probabilistic, unsupervised, bag-of-word model. In an iterative process LDA seeks to identify a set of topics which best differentiates the documents in a corpus. Then it assigns a probability to every term that it belongs to a certain topic in all the documents (Koblet and Purves, 2020). Thereby, it ignores semantics and word order, instead, focusing on statistical measures such as word occurrences and co-occurrences within a document (Murel and Kavlakoglu, 2024). The output of topic models is sensitive to a various range of input parameters such as the number of topics (Koblet and Purves, 2020). To determine the best number of topics a coherence and exclusivity analysis can be conducted which captures how qualitatively good a human perceives the generated topics. The coherence is a metric that indicates how interpretable a topic is by a human and the exclusivity defines how exclusive the top words are within a topic, which means that the higher the score, the less likely the words appear in other topics (Pandur, Dobša and Kronegger, 2020). The semantic coherence of a topic decreases as the number of topics increases while the exclusivity of words in topics increases with

the number of topics. Therefore, there is a trade-off between the two of them (Olsson, 2021). The exclusivity and coherence scores can be plotted for various number of topics until the optimum is found.

2.3 Natural Language Processing and Urban Space

Natural language processing is also increasingly used in urban research to answer questions that could not have been answered with traditional methods (Cai, 2021). In a systematic literature review Cai, p. (2021, p. 5) found that NLP methods in the urban context are mainly applied in five areas: "urban governance and management, public health, land use and functional zones, mobility and urban designs". Researchers most often use social media data or online platforms such as X (formerly known as Twitter), Instagram, Craigslist or Yelp to study the urban space and gentrification in different ways: Analysing online property listings in regard to neighbourhood names (McKenzie *et al.*, 2018), reconstructing the story of revitalisation in a small district based on a small corpus of different text genres from the study area (Luria and Campos, 2022), nowcasting gentrification based on Airbnb data (Jain *et al.*, 2021), analyse Yelp reviews to characterise neighbourhoods and neighbourhood change (Olson *et al.*, 2021), measuring policy debates and sentiment of a re-growth strategy in Leipzig with Twitter (Chen, Silva and Reis, 2021) and analysing the patterns and perception of gentrification in the United States based on Twitter data (Zenebe et al., 2018). In the following, I will briefly describe and summarise each of these studies and their methods to show how broadly NLP tools can be applied to analyse gentrification in the urban space.

In their paper "Greening a Post-Industrial City: Applying Keyword Extractor Method to Monitor a Fast-Changing Environmental Narrative", Luria and Campos (2022) compare the efficiency of the keyword extractor YAKE! compared to an expert reading of the same texts to study the narratives around the Canal District in Worcester, Massachusetts. Thereby, they gather a small corpus of 26 texts from both the present and past, newspaper articles to poems that describe the Canal Districts to study how the district transformed "from being spoken of as a 'dump' to 'the new IT town'" and how "language and its narratives impacts the city and the markets" (Luria and Campos, 2022, p. 110). Before they automatically extracted the keywords with YAKE! they pre-processed their corpus by transforming images to text so that they can be read by YAKE! as well as by removing noisy information to ensure the quality of the extracted data. Afterwards, they applied the keyword extractor to grasp the relevant information in the text. To be able to estimate the quality of the computationally generated results they also created a gold standard dataset where one of the authors manually identified keywords. However, this process proved to be challenging and subjective as well. In the end they compared the extracted keywords from YAKE! against the gold standard set for each text and marked down whether the automated results were good, sufficient or insufficient. Most results turned out to be good and therefore the authors conclude that YAKE! is a suitable tool to quickly produce a picture and show different narratives of a place. Most of all, they value the ability of YAKE! to generate word clouds with the extracted keywords that allow to grasp the content of a text at a glance. However, the authors also identify a few challenges with YAKE! which are that the keywords extracted are often repetitive and as it is a purely statistical tool, sometimes important words are omitted that crucially change the meaning of a text which leads to misinterpretation. They suggest applying a topic modelling algorithm to diversify the set of keywords as well as call out to continuously improve the tool.

Zenebe et al. (2018) applied a different approach to study gentrification. In their study they analysed the sentiment and perception of gentrification in the US based on twitter data by collecting all tweets containing the hashtag *gentrification* from the United States between 2015 and 2017. In their analysis they studied the co-existing hashtags, the authorship of tweets, compared the trend of the tweets with

trends in population growth and conducted a sentiment analysis of the content of the tweets. In the sentiment analysis Zenebe et al. (2018) automatically classified each word in a tweet as either objective, negative or positive then identified the target to what the sentiment refers. Then, they aggregated the overall sentiment in a tweet: If there were more positive than negative connotated words, the text was classified as "positive", vice versa the text was classified as "negative". If there were the same amount of positive and negative connotated words the text is ambivalent and if no sentiment terms were detected, the text counts as neutral. To show the results they used word clouds with the most shared negative and positive sentiments. They found that many tweets referring to gentrification related to an increase in population and thus demand for housing. Furthermore, they found that most tweets about gentrification were neutral but that tweets with negative sentiment were shared more often than tweets with positive sentiment. The authors summarise, that using big data such as twitter can reveal important insights and trends for decision making. However, they see a limitation of their studies in a possible inaccuracy of the contents due to short lengths of the text, informality of language and the credibility of the sources (Zenebe, Brown-Robertson and Mayo, 2018).

The paper of Chen et al. (2021) also conducts a sentiment analysis of twitter data but on a more specific case and shorter time period, namely the re-growth strategy of Leipzig 2030. To do this they gather both relevant tweets as well as news articles containing the keywords "Leipzig 2030" and "INSEK" as well as all the tweets geo-tagged in Leipzig for a period before and after the policy implementation. Thereby, they calculate the policy related tweets compared to all the tweets to measure the attention on the topic, compare the high-frequency keywords in both policy related and unrelated tweets as well as conduct a sentiment analysis. They found that attention on the topic is the highest right before and after the implementation but in general must conclude that the attention on the topic was rather low in the public. Furthermore, they could find a sentiment change from slightly more negative before the implementation to slightly more positive after the implementation. Nonetheless, they conclude that big data is an invaluable data source that allows to analyse the sentiment and popularity of a certain topic. Especially for public policy they claim that it is important to study the public opinion, as it can strongly influence the outcome and effectiveness of a policy (Chen, Silva and Reis, 2021).

McKenzie et al. (2018) use rental listings from Craigslist to identify neighbourhood names in the cities of Washington, Seattle and Montreal. They do this by using both natural language tools such as extracting n-grams and frequency of word occurrence in their corpus as well as machine learning approaches such as a random forest model to identify neighbourhood names within the n-gram dataset. Thereby, they find all craigslist rental listings for each of the cities in a certain time period and clean the data by removing duplicate and non-geotagged entries. They find that the approaches are suitable for identifying neighbourhood names in urban settings. However, the results show that in bigger, more dominant and popular neighbourhoods the probability of identifying the correct names in the n-grams is higher than in in smaller neighbourhood names. Furthermore, they find that there are some false positive results where names were identified that are not correct. Most often these were the names of landmarks, train stations or companies as they were mentioned in a rational way in the listings (e.g., apartment is close to xy). There were also false negative results in the model output which means that neighbourhoods were not properly identified. These were mostly neighbourhoods with a large share of residents living below the poverty line or with a high percentage of public housing. The authors explain this finding that either there were too few listings in these areas so that they were removed or that the names of the neighbourhoods were not mentioned in the listings. This points, so the authors, to a bias in the datasets as most listings are created by people from higher socio-economic groups such as real estate agents or property owners (McKenzie *et al.*, 2018). Despite this bias in the data the authors conclude that this bottom-up, data-driven approach is suitable to identify existing as well as potentially alternative neighbourhood names.

The paper “reading the city through its neighbourhoods: Deep text embeddings of Yelp reviews as a basis for determining similarity and change” by Olson et al. (2021) goes in the same direction as they try to identify neighbourhood boundaries as well as characterise a neighbourhood based on Yelp businesses review data. The study focuses on Toronto in a 10 years’ time frame and uses natural language processing tools such as word n-grams, frequency of words and topic modelling as well as a vector model to characterise the neighbourhoods. With the vector model they seek to show the similarity between neighbourhoods as well as how each neighbourhood changes over time. The authors find that Yelp data is suitable to identify both neighbourhood boundaries as well as neighbourhood change. However, one of the limitations the authors see is the bias in social media data, where few people create much of the content (Olson *et al.*, 2021).

Similarly to this, Jain et al. (2021) analyse whether it is possible and if so, how well, to nowcast gentrification based on Airbnb data. The importance for this, so the authors is, that in order to mitigate the negative effects of gentrification it is important to have up-to-date and readily available data. Because this is not possible with traditional surveys that are expensive and only take place every few years, real-time, accurate online data should be used. For their study they first conduct a literature review and find that measuring urban change with alternative data is a growing field of interest, enabled through the rapid growth of online platforms and social media. In a second step they construct a gentrification score for each neighbourhood in New York, Los Angeles and London based on changes in four socio-economic measures that have been identified as relevant in studying gentrification: age, education, housing affordability and income for the two time-windows 1998-2002 and 2013-2017 (Jain *et al.*, 2021). Based on this gentrification score they then select the disadvantaged neighbourhoods in the three cities as they are fast changing and mostly discussed in regard to gentrification. For the time between 2013-2017 they then collect both structured (number of listings, average price, average number of bedrooms) as well as unstructured (reviews and descriptive text) Airbnb data. For the unstructured data they extract the review lengths, location words, sentiment und topics in the individual listings with NLP tools. For the main analysis the authors examine the correlation between the gentrification score and the structured and unstructured Airbnb data, compare the frequency of location words for gentrifying and non-gentrifying neighbourhoods and use a regression model (simple linear and random forest model) to predict a gentrification score based on the Airbnb data. They find that there is a high correlation between gentrification and Airbnb data and certain location words appear significantly more often in gentrifying neighbourhoods than in places where gentrification is not occurring. Furthermore, their results show that unstructured features in form of texts have more explanatory power than their structured counterparts, signalling the importance of text analysis. However, as in previous studies Jain et al. (2021) see the biggest limitations of their approach in the bias of the data towards the predominantly affluent, educated and younger users of the platform as well as the limited timeframe since Airbnb has only substantially grown since 2013.

The short literature review of six recent studies that apply NLP methods to the urban context show that the tools are used for very different purposes and topics: from characterising neighbourhoods, to measuring change and analyse sentiment. This indicates that it is an active research field with a broad variety of methods and topics. The above papers are all situated in the field of Geographic Information Science and are very useful for me to develop the methodology for this thesis. Furthermore, they all highlight limitations of these methods of which one should be aware of.

3. Methods

3.1 Tools

For the analysis I mainly used Python (version 5.4.3) in the Spyder environment as it contains many built in Natural Language Processing (NLP) libraries such as NLTK or SciKit. Python was used to pre-process the corpus and perform all the NLP methods on it as well as generate word clouds for the collocation results. The map visualization for the occurrence analysis was done in QGIS.

3.2 Research Area – A Short History of Zurich North

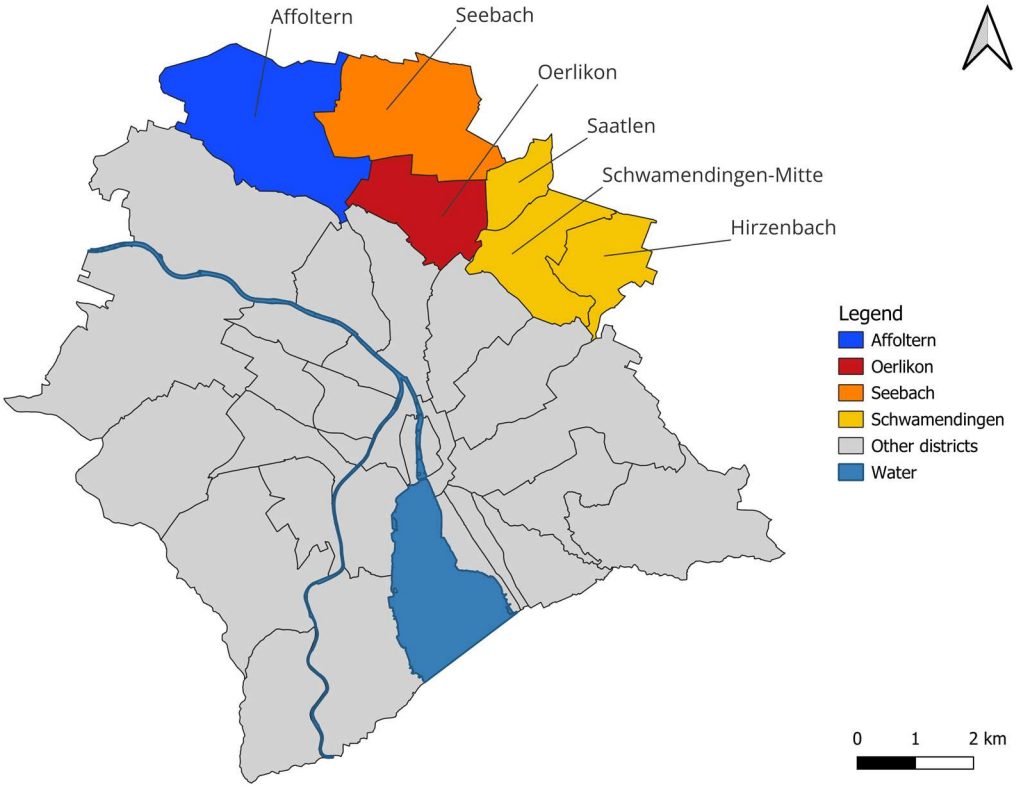

Figure 1: Oversight of the study area with the four highlighted neighbourhoods Affoltern, Oerlikon, Schwamendingen and Seebach.

The area Zurich North is located in the north of the city of Zurich, Switzerland. The region consists of two different districts, so-called "Kreise", four different, equally sized neighbourhoods and seven quarters. Kreis 11 is divided in the neighbourhoods Affoltern, Oerlikon and Seebach. Kreis 12 comprises the neighbourhood Schwamendingen which is again sub-divided in three quarters Schwamendingen-Mitte, Saatlen and Hirzenbach (Figure 1). However, the three quarters are mostly used for official units which is why I will mostly focus on Schwamendingen in total and not the three individual quarters. The two districts are the newest in Zurich and have been officially incorporated in the city in 1934 (City of Zurich, 2023b). In 1855 the train station in Oerlikon was built which triggered an economic boom in the second half of the 19th century. Today, the train station in Oerlikon is the 8th biggest trains station in Switzerland passed on passenger volume per day (SBB, 2024). With the train station many industrial firms such as the machine factory Oerlikon (MFO) or the Oerlikon-Bührle settled in Oerlikon in the following years due to its good position at a main transport node (both train and car). Many workers and their families were attracted by these factory jobs. This caused a great need for housing in the surrounding area of Oerlikon, leading to a steady increase of its population (Figure 2). During this time a lot of housing units were built for the workers, often by housing cooperatives. The industrial history of Zurich North is still reflected in the high prevalence of housing cooperatives in these neighbourhoods

today. After a short demographic decline between 1960 and 1990, the population has since been steadily increasing, transforming Zurich North yet again (Figure 2).

Oerlikon is the centre of the northern districts in Zurich with many shops, the local authority and cultural centres such as the Hallenstadion, the local indoor pool or the Theater 11. From the northern districts, Oerlikon is also the only one with an industrial history. By 1988 most of the industry and

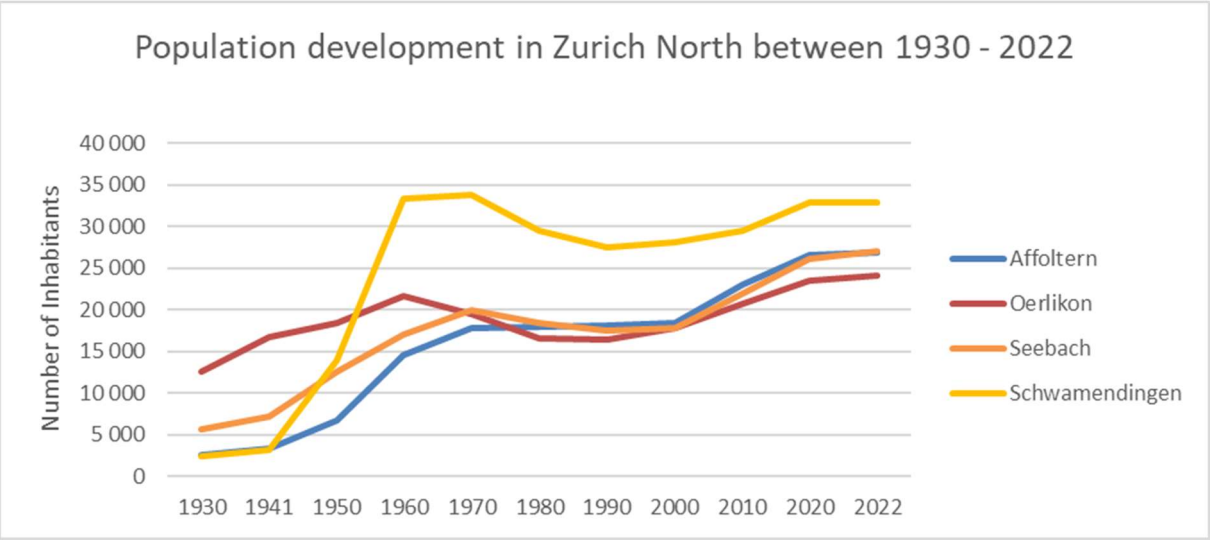

Figure 2: Population development in Zurich North between 1930 - 2022 split in the four neighbourhoods. Of note is the high population growth between 1940 and 1960 and the steady increase since 1990.

production factories have moved away from Oerlikon and thus a plan (Zentrum Zürich Nord, ZZN) was created to redevelop the former industrial areas into living and working space. The creation of Neu-Oerlikon with an area of 60 hectares was one of the biggest redevelopment projects in Switzerland and created ca. 12'000 working spaces and living room for 5'000 people (City of Zurich, 2023b). In the last twenty years many high rising buildings were developed close to the train station such as the Franklin- or Andreasturm, redefining the image of Oerlikon. Affoltern lies to the west of Oerlikon and originally consisted of two individual villages. With the industrial revolution many workers moved to the neighbourhood and there was an intensive settlement formation that formed the neighbourhood "Neuaffoltern". In the middle of the 1950s Neuaffoltern extensively expanded, and a lot of new housing was constructed as the population more than doubled (City of Zurich, 2023a). Schwamendingen has a very similar story than Affoltern and used to be a small farmer village before the industrialisation of Oerlikon. The neighbourhood saw the biggest population growth between the 1940s and the end of the 1950s, where the population sextupled from under 5'000 to over 30'000 within twenty years (Figure 2). During this time that most of the housing was built based on the ideals of the "garden city" (Widmer and Kübler, 2014; City of Zurich, 2023c). Currently many of these buildings are being replaced with new builds, massively transforming the neighbourhood. Also, Seebach grew substantially due to the industrialisation of Oerlikon at the end of the 19th century. Until the 1950s Seebach was oriented towards Oerlikon and did not really have any local businesses or industry. This changed with the construction of the national television studios in the former wetlands of the Leutschenbach. The Leutschenbach is located in the south of Seebach bordering Schwamendingen and is one of the biggest development areas in Zurich. The re-development of the former industrial site commenced in the early 2000s and is still ongoing, adding thousands of additional living room and office spaces (City of Zurich, 2023d).

3.3 Data

3.3.1 Building a Corpus of Newspaper Articles

The corpus is built by using the Swissdox@LiRi tool which allows to retrieve large quantities of Swiss media data for research purposes (Swissdox@LiRI, 2024). The Swissdox library contains approximately 24 million published media articles from many different local, regional and national Swiss media sources. While most articles are from the last 25 years, it also includes articles that date back to the beginning of 1900. To gather the corpus, I used the provided code by Swissdox to start a query with an API. Thereby, I searched articles from 01.01.1850 until the 01.03.2024 in German. I decided to restrict the search to only German articles since Zurich is in the German speaking part of Switzerland and therefore, I assume that most relevant articles will be in German. Additionally, having different languages in the same article could lead to further complications.

To select the articles, I created a gazetteer that contains all the street names in Zurich North, based on the "Strassennamenverzeichnis" of the city of Zurich, which is updated daily (data accessed on 26.01.2024) (City of Zurich, 2024i). The street names gazetteer was then manually annotated with important places and building projects in the neighbourhoods mentioned in the "Quartierspiegel" of the six districts. The "Quartierspiegel" portrays all neighbourhoods in Zurich and is annually updated by the city of Zurich. It contains historic information as well as recent events and developments in the districts (City of Zurich, 2024h). Additionally, three parks that were not in the street name directory (Gustav-Ammann-Park, Leutschenpark and Oerliker Park) were added based on the work of Sophie Sturzenegger who manually created a list of all the parks and places in Zurich for her master thesis (Sturzenegger, 2022). Furthermore, the statistical neighbourhoods used by the Stadt Zurich were added to the gazetteer. In the end, the Zurich North gazetteer contained a total of 561 locations which were entered in the Swissdox API.

Since street names are not unique in Switzerland (e.g. Bahnhofsstrasse), a search in Swissdox with only the street names could retrieve many articles that do not relate to Zurich North. To avoid such mix-ups, I added a must criterion to the query so that only those articles are extracted that contain at least one of the 561 locations as well as one of the following neighbourhood names (Oerlikon, Schwamendingen, Affoltern, Seebach, Saatlen, Hirzenbach), Zürich, Kreis 11 (district 11) or Kreis 12 (district 12). Thereby, I assume that an article that contains a street name also includes a more general regional term that allows readers to place the mentioned location. A short manual evaluation of a few newspaper articles supports this assumption.

With this search query, a total of 59'261 articles was found on Swissdox@LiRi, whereby, 52'358 had a unique content ID and header (6'903 duplicate articles) (Table 1). After a first analysis of the frequency results for each of the locations in the Gazetteers, I was surprised to find that most results were found for Neumarkt. After reading some of the articles containing Neumarkt, it became apparent that most of the articles were not about the Neumarkt in Oerlikon but the Neumarkt in the inner city. This is why I decided, to remove all the articles that contained only Neumarkt from the corpus. The first article that was retrieved from the API search is from 1913. Looking at the distribution of the retrieved articles, it became apparent that there are only very few articles before 1991 (between 1913 and 1990 only 778 articles were found compared to 9'707 articles between 2016 and 2020 alone) (Table 2). Even though it would have been interesting to also analyse older data, I decided to focus only on articles that appeared after 1990 to avoid large differences in dataset sizes. In total, 44'172 articles were used for the analysis containing over 19 million tokens (Table 1)

Looking at the source of the newspaper articles, Swissdox@LiRi retrieved articles from 151 different sources. The articles were gathered from both print and online, daily and weekly as well as regional and national sources. To closer look at the distribution of the sources, I first grouped newspapers that were listed as different sources but belong to the same newspaper (for example NZZ, NZZ Folio, NZZ

am Sonntag and nzz.ch were grouped under NZZ). In the end, this resulted in 78 different newspaper sources. Most articles stem from the two biggest newspapers in Switzerland, both located in Zurich: Tages-Anzeiger (10'610 articles) and Neue Zürcher Zeitung (NZZ, 9'116 articles). Both newspapers are published in the whole of Switzerland but contain a subsection about Zurich. With just over 2'000 articles, the Zürcher Unterländer, a regional newspaper has the third most articles. The other newspaper sources in the top ten are: Blick, Basler Zeitung, Der Landbote, Zürichsee-Zeitung, Limmattaler Zeitung, 20min and Der Bund. Out of the top ten, four sources are regional newspapers from within the canton of Zurich and the rest are national newspapers based in different cities. Unfortunately, the most important local newspaper in Zurich North called Zürich Nord is missing from the dataset. After a short inquiry, they let me know that their archive is not yet available digitally. Furthermore, it must be noted that many of the newspaper sources belong to the same parent company such as CH Media, Ringier or NZZ which could lead to a certain bias if articles are being shared among the newspapers. Nonetheless, the corpus provides a broad spectrum of articles that is suitable for an NLP analysis.

Table 1: Summary of the number of articles and tokens that resulted from the Swissdox@LiRi search request and those that were excluded from the further analysis. In total 44'950 articles that included more than 19 million tokens were found

Type of Articles	Number of Articles
Total articles of initial results	59'261
# of duplicate articles	6'903
# of "Neumarkt" articles	7'408
# of articles before 1991	778
Total number of articles	**44'172**
Total number of tokens	**19'295'208**

Table 2: Number of retrieved articles for each time period. Because there were only very few articles extracted from before 1991, those were excluded from the analysis.

Period	Number of Articles	Number of Tokens
1913 - 1990	778	488'047
1991 - 1995	937	399'013
1996 - 2000	4'559	1'869'577
2001 - 2005	6'192	2'042'139
2006 - 2010	8'940	2'169'241
2011 - 2015	7'441	2'067'148
2016 - 2020	9'707	5'056'241
2021 - 2024	6'396	5'691'849
Total	**44'950**	**19'783'255**

3.3.2 Building Activity in Zurich

To compare the NLP results with the actual building activity during different time periods, I used the Swiss Federal Housing and Building Register (GWR) which can be downloaded online (FSO, 2024). The GWR is a point-data set that contains the coordinates and additional information about each building in Switzerland. Additional information includes among other things the year of construction, usage, number of stories, number of apartments, building volume and heating type. Besides containing the current building stock in Switzerland, the GWR also entails planned projects as well as a historization of former buildings. This allows to track the building activities in previous years. The GWR is updated on a regular basis. In this thesis I used data downloaded on the 15.12.2023 which means that buildings that were constructed in 2024 are missing in the analysis.

3.3.3 Data Periods

In this thesis I compare the newspaper discourse on gentrification and neighbourhood uplifting across time as well as overall. After careful consideration, I decided on five-year time periods as this balances a reasonably high temporal resolution as well as time feasibility restrictions. Within five years a neighbourhood can change significantly especially when whole areas are being developed, which makes it an interesting time span to analyse. For the analysis, I grouped both the newspaper corpus as well as the GWR data in 7 different time periods based on the publishing date and year of construction. The different time periods are: 1991 – 1995, 1996 – 2000, 2001 – 2005, 2006 – 2010, 2011 – 2015, 2016 – 2020, 2021 – 2024. It is important to note that the last period only consists of three full years and two months due to the point of data extraction. Even though this time period is much shorter than the others, there was enough data available both in the corpus as well as the GWR to enable an analysis. Furthermore, since the housing shortage became more acute in recent years, I believe it is important to include this period in the analysis as well. Besides applying the different methods for each period, I also looked at the corpus as a whole, including all articles and buildings constructed between 1991 and 2024.

3.4 Pre-Processing the Data

3.4.1 Removal of Duplicates and Text Snippets

The first step in pre-processing the data was to remove special characters such as question marks, asterisks or brackets in the text as well as HTML code with a regex expression. The HTML code such as e.g. “<tx><ld><p>” are a peculiarity from generating the corpus with Swissdox@LiRi AP. Since neither the special characters nor the HTML code holds any relevant meaning but appear in high frequencies, they were removed to reduce noise in the data and avoid later extraction with the statistical NLP methods. The second step of pre-processing I conducted was to remove duplicate articles. I defined duplicates as those articles that have the same content ID extracted by Swissdox@LiRi. Additionally, I also intended to remove those articles that have the same newspaper headers. However, after a short evaluation of different articles, I decided against it, since some articles that had the same header contained different text and were published in different years. Since NLP extracts words based on statistical frequency, duplicated articles would heavily influence the outcomes. Thus, removing duplicates is important to reduce bias in the outcomes.

3.4.2 Tokenisation, Removal of Stopwords and Normalisation

The next step in the data pre-processing was to tokenise the text to facilitate the removal of stopwords and normalisation. Thereby, I used a pre-trained German tokenisation model from the NLTK toolkit in python to first tokenize each article into sentences and then each sentence into words. After the tokenisation I removed all stopwords from the corpus. Many stopword-lists already exist online also for the German language. They differ both in content and length and therefore the choice of stopword list

can have a significant impact on the results. After comparing different stopword-lists, I chose to use a German stopword list called "stopword-de" I found on Github that contains 620 words in total (Stopwords Iso, 2020). It is according to the website the "most comprehensive collection of stopwords for the German language" (Stopwords Iso, 2020) and is thus preferable over the already built in stopword list in the NLTK toolkit which only includes 232 stopwords.

The last pre-processing step in this analysis was the normalisation of all text in the corpus. Thereby, I decided to prefer lemmatising over stemming as the reduced word can always be found in a dictionary and is thus easier to interpret in the results. In Python, many different lemmatisation packages exist such as in NLTK or Spacy. However, most of them have initially been developed for the English language and not all of them offer lemmatisation for German text (Wartena, 2023). Therefore, I chose to use the Hannover Tagger (HanTa), which was initially built for the German language with the main goal of being "easier to use" (Wartena, 2023, p. 4). HanTa contains a morphological analyser, a part of speech tagger and a lemma generator. It is based on traditional machine learning techniques and gives state of the art result for lemmatisation and PoS tagging for the German language (Wartena, 2023).

I decided against using contraction expansion and spelling correction as recommended by Sarkar (2016). The reasons for this are that contractions are less common in German than they are in English and that newspaper articles tend to be revised by an editor before publishing. Thus, I assume that the data is already in good quality, compared to for example social media data.

3.4.3 Article Categorisation

In my analysis I focus on five different dimensions of urban transformation and the gentrification process which are at the core of my thesis. These five dimensions are *Gentrifizierung* (gentrification), *Aufwertung* (uplifting), *Neubau* (new build), *Sanierung* (renovation) and *Verdrängung* (displacement). I picked these keywords based on the literature review on urban change and personal interest to find out how their usage varies over time and space in the corpus and how they match with the building activity. The five keywords can be divided into 3 different categories: objective newspaper coverage (*Neubau*, *Sanierung*), positive coverage (*Aufwertung*) and negative coverage (*Gentrifizierung*, *Verdrängung*).

To tag each article with the corresponding topic and neighbourhood, I added additional columns for each of the five topics as well as the four neighbourhoods (Affoltern, Oerlikon, Schwamendingen, Seebach) to the corpus. For the thematic categorisation, I looped through all the articles and checked if the article contained either the noun or the verb of the corresponding keyword. If it did, the column was appended with *ja* (yes), if none of the words was found the column was appended with *nein* (no). This step was repeated for all the five topics. Articles containing new build and renovations occur most often in the corpus, with over 3'000 articles. In a third place with more than 1'000 articles comes uplifting while displacement occurs only in roughly 500 and gentrification only in roughly 100 articles (Table 3). Articles that contain either one of the keywords are considered to be thematically relevant and make up the "topic corpus" used in the collocation analysis. To tag the articles according to the neighbourhood they are covering, I preceded in a similar way. Thereby, I labelled each entry in the gazetteer with the corresponding district(s). With over 14'000 articles, Oerlikon and Seebach occur the most, followed by Schwamendingen with over 10'000 articles and Affoltern with more than 8'000 articles (Table 4). The thematical and spatial allocation of the articles enables an efficient way to compare the newspaper coverage per district and topic over time, the creation of sub-corpora as well as the selection of interesting articles for close reading. The one-off tagging of the articles saved a lot of time in later analyses and data inspection as it allowed to filter the corresponding sub-corpus directly instead of looping through all articles again.

Table 3: Number of articles that contain the following keywords: Gentrification (German: Gentrifizierung, gentrifizieren, gentrifiziert), uplifting (Aufwertung, aufwerten, aufgewertet), new build (German: Neubau, neubauen, neugebaut), renovation (German: Sanierung, sanieren, saniert), displacement (German: Verdrängung, verdrängen, verdrängt).

Keyword	Articles	Frequency	# Tokens
Gentrifizierung (Gentrification)	97	103	64’137
Aufwertung (Uplifting)	1’068	1’584	828’666
Neubau (New building)	3’813	6’279	2’693’450
Sanierung (Renovation)	3’619	6’902	3'094’716
Verdrängung (displacement)	494	529	348’917
Total	**9’091**	**15’397**	**7’029’886**

Table 4: Number of articles per district. Noteworthy is that most articles are about Seebach and Oerlikon and the least about Affoltern. With this method 6’500 articles could not be tagged at all. Since there are articles that cover more than one neighbourhood, the total number of articles exceeds the total number of articles in the corpus.

Neighbourhood	Articles
Affoltern	8’335
Oerlikon	14’336
Schwamendingen	10’839
Seebach	14’817
Total	**48’327**

3.5 Data Analysis

This thesis applies three different NLP methods: Frequency and co-occurrence, collocation and topic modelling to a corpus of German newspaper articles about Zurich North between 1991 to 2024. The main goals are to explore how these three NLP tools can be used to reveal newspaper discourses on gentrification and how analyse how newspapers reported on this issue in the past 30 years. In the following, I will explain how each method was implemented which was then applied on the whole corpus as well as seven different time sub-corpora. The different time sub-corpora were built by applying a filter mask to the corpus, filtering based on the publishing year of the article.

3.5.1 Frequency, Co-Occurrence and Comparison with the GWR

Frequency

To calculate the frequency, or occurrence as I also refer to it in this thesis, of each of the 561 gazetteer entries, a feature matrix was created. A feature matrix is a matrix that includes all the articles as rows and all the words that occur in the corpus as column names (Figure 3). For each article, the matrix depicts how often a word occurs. Thereby, different measures exist to build such a matrix. The scikit-learn module "feature extraction" offers two different vectorizers: the CountVectorizer and the TfidfVectorizer. Based on these two vectorizers, three different feature matrix measures are available: binary, frequency and TF-IDF. The binary feature matrix depicts for each article if a word occurs or not (1 if the word occurs, 0 if it does not). The frequency feature matrix counts how often the word occurs in each article and the TF-IDF matrix calculates the TF-IDF score for each word. After building the feature matrix, the values for each of the words are summed up to generate a frequency score per word.

To decide the best suited feature matrix measure, I applied all three of them on the corpus and compared the results. The binary measure works fine and is suitable for the approach. However, locations that appear often in the corpus in general appeared the most in all the co-occurrence maps, potentially overshadowing other, more interesting places with lower frequencies. The frequency measure for the feature matrix on the other hand is not suited for the analysis since articles that contain the same location name multiple times would be counted more than once. This would introduce a bias in the frequency data and overestimate the importance of certain locations. The results for the TF-IDF are weighted by document frequency and thus less biased towards places that generally occur a lot in the corpus. This is the reason, why I chose to use the TF-IDF weighted results for my analysis. However, one drawback to this approach is that it becomes more difficult to interpret and verify the results by filtering the data.

Index	integrieren	inteiligent	intelligent	intelligenz	interessant	interesse	interessiert	investition
0	0	1	2	1	1	3	2	0
1	1	0	0	0	0	0	1	5
2	0	0	0	0	0	0	0	0

Figure 3: Example of a Frequency Feature Matrix based on the first three articles in my corpus. For each of the articles (rows), the occurrence of a word (columns) can be seen. For example, in the second article, the world "investition" occurs 5 times.

Using this feature matrix approach is an elegant and time efficient way to calculate the frequency of certain words in a corpus and is preferable to an implementation with nested for-loops which takes a long time to compute. However, there are a few issues that one needs to be aware of. First of all, a feature matrix splits up all tokens in the corpus. This can become a problem for locations that are consist of multiple words such as Hunziker Areal or Neu-Oerlikon. Before building the feature matrix,

the text was therefore pre-prepared again to take this into account. In the corpus, all street names that consist of multiple words were replaced by versions that were joined by an underscore. This prevented the feature matrix from splitting up the individual tokens (e.g. Hunziker_Areal instead of Hunziker Areal). Another, more general problem with counting frequencies with a feature matrix are different spellings of a location (e.g. Hunziker Areal vs. Hunziker-Areal vs. Hunzikerareal), spelling errors or different lemmatisations during the pre-processing of street names (e.g. Kreuzwiese instead of Kreuzwiesen). There are different options to counter this problem. One possibility to address this issue would be to use a fuzzy matching approach when summing up the results of the feature matrix. Another option would be to extract all location names with a Part-of-Speech (POS) tagging, then normalise the different spellings to one basic form and lastly preserve this spelling form during the pre-processing. However, both the fuzzy matching and the location extraction with a POS tagger are not trivial to implement and potentially very time-consuming. To still avoid the issue of different spellings, I looked more closely at those locations that did not yield any results in the frequency analysis and looked for potentially different kinds of spellings. Furthermore, I applied the same pre-processing steps that I performed on the corpus to the location gazetteer and repeated the frequency analysis. In the end, I selected the spelling(s) that yielded results and summed them up.

Co-Occurrence

The co-occurrence in this thesis is defined as the co-occurrence of a place name in the gazetteer with each of the five keywords in an article. Thereby, it does not matter how close the two words are in an article, but just whether they occur together in the article. To calculate the co-occurrence of the locations in the gazetteer, a very similar process than the one computing the frequency was used. First, a feature matrix with the TF-IDF score was built over the whole corpus. In a second step, articles that contain both the location and the keywords are identified and their TF-IDF values are multiplied. Lastly, the multiplied TF-IDF scores are summed to receive one co-occurrence score for each location.

Comparison of Frequency and Co-Occurrence with the building activity in Zurich

The calculated frequency and co-occurrence scores per location were first categorised in one of 6 criteria based on their values. The categories were defined after an evaluation of all the frequency and co-occurrence scores to best reflect their distribution. Since most locations had relatively low scores, the lower scores are categorised more finely than the high scores. For a better comparability, the same categorisation was used for all time periods and co-occurrences. The following categories were used in the analysis: 1 – 5 occurrences, 6 – 10 occurrences, 11 – 20 occurrences, 21 – 50 occurrences, 51 – 100 occurrences and more than 100 occurrences.

To create maps for the different topics and time periods, the categorised data per location was then uploaded to QGIS where it was joined with the street network data of the city of Zurich (City of Zurich, 2024i) and a point dataset of the parks and POIs. The results of the frequency and co-occurrence analysis was then displayed by a colour gradient to highlight places with high (co-) occurrences. To visually compare the data with the actual building activity, the GWR point data set was added in the background in form of a heat map. Thereby, the heat map was weighted by the building volume to emphasise large projects with a higher volume over smaller buildings. This highlights the impact of large and especially high-rise buildings (that do not cover a lot of area but have a high volume) on the aesthetical look and feel of a neighbourhood. However, other weighing measures such as building area, number of stories, number of flats, population or project costs could also have been used, highlighting a different aspect. For each of the seven time periods as well as over the whole period, maps were generated consisting of the frequency of the locations as well as the co-occurrence with the different keywords.

To support the visual analysis, the relationship between newspaper coverage and building activity were also assessed statistically. For this, the construction activity had to be set in relation to the newspaper coverage. Since in my analysis there are both point features (squares, parks and POI) as well as linear features (streets), two different approaches had to be taken. For the point data, a hectare grid was created over the whole area, summing up the building volumes in each cell per period. In a second step, each point in the dataset was joined with the summed building volumes in the cell it is positioned. For the street data, a different approach had to be taken since a street does not fall in just one cell of a hectare grid. Instead, the GWR building data, weighted by volume, was rasterised in a 100x100 grid and the streets were buffered with a 20m buffer. Then, a zonal statistic was carried out, calculating the mean building volume value from all the cells a street passes. For each street and each time period the mean building volume was calculated and saved.

Finally, all the point data and linear features included both the frequencies and co-occurrences as well as the building volume constructed during each period. Based on this dataset a bivariate analysis was conducted by calculating the Spearman's rho. Spearman's rho was calculated for different subsections of the data and for each of the periods and topics. The subsections used for the analysis were: all locations, non-profit locations (e.g. housing cooperatives), for-profit locations, locations in Affoltern, locations in Oerlikon, locations in Schwamendingen and locations in Seebach. The locations based on the ownership was marked based on the dataset "Wohnbauten gemeinnütziger Bauträger" (Residential buildings of non-profit property developers) provided by the canton of Zurich (Canton of Zurich, 2024). A relation between the newspaper coverage and building activity is only statistically significant, if the p-value is below 0.05. The strength of the relationship can be evaluated based on the categorisation of Cohen (1992): strong (r values greater than 0.5), moderate (r values between 0.3 – 0.5) and weak (r values between 0 and 0.3). Thereby, positive values indicate a positive relationship while negative values indicate a negative relationship. I decided to use the Spearman's rho since it calculates the relation between two variables and does not require a normal distribution of the data (UZH, 2024).

3.5.2 Collocation

While the frequency and co-occurrence analyses reveal where newspapers talk about gentrification and urban transformation as well as in what context, a collocation analysis can give more insights on how newspaper talk about certain topics and how it changes over time.

In this thesis I used the BigramCollocationFinder from the NLTK toolkit, which has a collocation window span of two, to find the top twenty collocations for a set of nodes I defined. As an association measure I used the student t-test, since according to Brezina (2018) it balances the frequency and exclusivity. Further, I decided against applying a frequency filter in the collocation analysis, which would only display words that collocate with the node more than x times. Especially for words that appear very infrequent in the corpus, applying a frequency filter could lead to empty results.

For the collocation analysis, 56 nodes were defined based on literature research and theoretical interest that are connected to the study area, the five keywords and terms generally used in the context of city development (a list of all nodes can be found in Appendix A). For each of the nodes, a collocation analysis was conducted for nine different sub-corpora to see how the discourse changes over time and corpus. In a first step, the collocates were extracted for the whole corpus. In a second step, the collocates were extracted for the thematic corpus containing all articles that were tagged with at least one of the five keywords (gentrification, uplifting, new build, renovation and displacement). Comparing the whole corpus against the thematic corpus enables to examine whether a node is also used in different contexts besides urban transformation. Thereby, I expect that if the collocates in both corpora

are similar, the node is predominantly used in an urban development context. Otherwise, I assume that the node is used in other contexts as well. Thirdly, the collocates were computed for each of the seven time periods within the thematic sub-corpus to analyse if the discourse around the node in the context of urban change has changed over time.

To visualise the data, the collocates for each sub-corpus are displayed in a word cloud. Thereby, the collocates are weighted based on their collocation frequency: the more they collocate, the bigger they will be displayed in the word cloud. Furthermore, node, the sub-corpus and the frequency of the node in each sub-corpus is indicated in the title. If no collocates were found in a particular sub-corpus, the word cloud is empty.

3.5.3 Topic Modelling

For the topic modelling, a Latent Dirichlet Allocation (LDA) algorithm was applied to the five different topic sub-corpora (gentrification, uplifting, new build, renovation and displacement) to investigate in what thematical context they are used in. Building an LDA model requires a reasonably large dataset to generate good results and since some of the topic sub-corpora data sets are already relatively small (e.g. the gentrification corpus only contains 97 articles over the 30-year time span), the topic modelling analysis does not include a further breakdown by time-period. Topic modelling can take a long time to compute whereby it takes longer, the more tokens there are in a corpus as every term is assigned to one topic. This proved challenging as some of the sub-corpora contained a lot of tokens (Table 3) and the processing capacity of my personal computer was limited. To cut down computation time, I thus decided to reduce the number of tokens before using LDA. To reduce the number of tokens, only nouns were included in the topic analysis by using the POS tagger included in the HanTa module, filtering for all noun categories. The focus on nouns was chosen, given their importance in identifying key concepts and topics within a text. Further, I reduced the number of tokens by removing the 10% most frequent and the 5% rarest nouns, to filter out terms that are either too exclusive or too abundant, focusing on terms that are the most informative.

To find the best LDA model, a GridSearchCV was performed by varying different input hyperparameters such as number of topics (5 - 10), learning offset (1, 10, 100, 1000) and learning decay (0.6, 0.7). The grid search was aimed at maximizing the model's log-likelihood and minimizing perplexity, which serves as a measure of how well the model predicts the data. The best LDA model was then selected based on the grid search results and the model refitted on the entire dataset. The model was used to assign each document to its most likely topic and to identify the most prominent terms for each topic.

To visualize the topics identified by the LDA model, the top 20 words in each topic were extracted and plotted using horizontal bar charts. This visualization allows to grasp the identified topic and enables a comparison across the different topics. Furthermore, the associated weights of the top 20 words were plotted in the bar chart, indicating the importance of each word within the topic.

Besides implementing an LDA approach, I also experimented with a Large Language Model (LLM) that uses artificial intelligence to summarise topics within a large corpus. For this the open-source AI Llama was used to identify topics in the gentrification sub-corpus. In a first prompt, Llama was asked to summarise the ten main topics per article and return them in a list. With a second prompt, the list with the top ten words per articles was then clustered into twenty topics, containing the top twenty words per cluster. In a third prompt, Llama was asked to condense the list to a maximum of twenty words by merging similar topics into single words. Although the approach was quite promising it is also very resource and CPU-intensive already for the smallest of the sub-corpora with only 97 articles. Therefore, I decided to apply this method to the gentrification sub-corpus only.

4. Results

This chapter examines the results of the three different Natural Language Processing (NLP) methods applied to the corpus. In the first subchapter the results from the frequency and co-occurrence analysis are described and compared to the construction activity. It starts with a broad general overview of the study area and which places have been mentioned in the newspaper corpus and which have not. Furthermore, this part looks at the distribution of how often locations occur in the corpus in general. In a second part, the results from each of the seven time periods as well as the results over the whole corpus are discussed. Thereby, the analysis always follows the same structure: first the maps showing the co-occurrence for each location with the keywords as well as the building activity are depicted (an enlarged example can be found in Appendix D). Second, an overview of the building activity during that period is given. Third, the places with the top five frequencies and co-occurrences are addressed and, in some cases, complemented with close reading. Special attention will thereby be given to the articles containing gentrification as it is the main focus of this thesis. Lastly, the correlation between newspaper coverage on gentrification and urban transformation with the actual construction activity is evaluated as well as locations where there is a gap between coverage and construction activity. I decided to focus on the top five results to enable a more detailed discussion of the most relevant locations during a period. Nonetheless, it would have been exciting to look at other locations as well, but the sheer number of locations would have been difficult to manage within this thesis.

In the second subchapter the results of the collocation analysis are analysed in detail. Thereby, the focus lies mostly on the collocation results of the different neighbourhoods as well as the five keywords to find out more about the context they are used in and how this may have changed over the past 30 years. Additionally, the most interesting results from words of interest are discussed in detail. Not further examined will be collocation analyses of words that did not yield any results or results that did not reveal anything of interest. The collocation results will be analysed based on the generated word clouds and supported with some close reading.

In the last subchapter the results of the topic modelling are reviewed for each of the thematic sub-corpora. Based on the extracted keywords, each topic is assigned a main category, if possible. For each of the identified main categories, it is discussed how it might relate to the sub-corpus and interesting keywords are elaborated on. Furthermore, the similarity between the different main categories as well as the different sub-corpora are looked out. However, the output of the topic modelling is not the main focus of this thesis, and the focus lies more on what additional value it can bring besides the collocation and frequency / co-occurrence analysis.

4.1 Frequency and Co-Occurrence

4.1.1 Occurrence of Locations in the Corpus

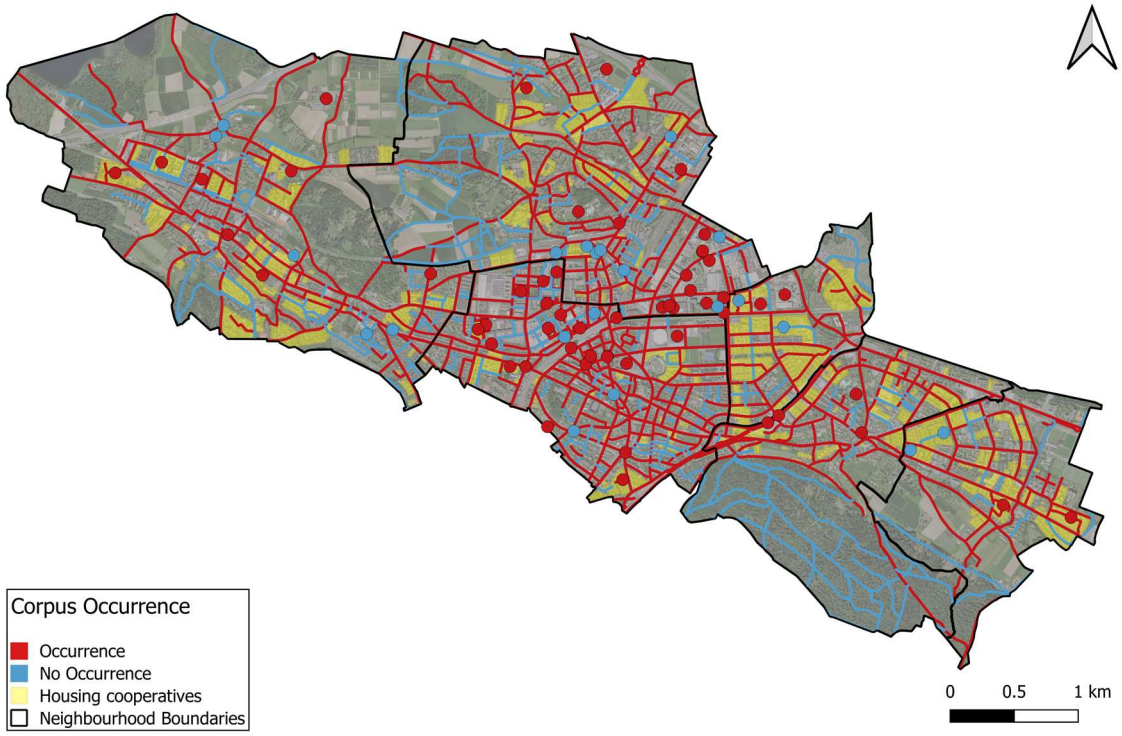

Figure 4: Map of the whole street network, parks and POI as well as social housing estates in Zurich North grouped by their occurrence in the newspaper corpus. Noticeable is that most of the locations were mentioned at least once in the corpus.

Figure 4 displays the street network, parks and POI in Zurich North and shows which locations occur in the newspaper corpus, and which do not. Out of the 561 locations in the gazetteer, 293 occurred at least once in the corpus during the analysed time span while 268 locations were never mentioned. One observation is that streets that pass through a forest or agricultural fields are rarely mentioned in the news corpus. This can be especially observed in the south of Schwamendingen as well as the west of Seebach. A possible reason for this could be that not many people live close to these streets and that there is generally not much going on in these areas that is newsworthy. Furthermore, it could be that these street names are less relevant and known since they are predominantly used for leisure activities or agricultural and forestry purposes, where the name of the streets is needed less for orientation. Another striking observation is that many of the shorter streets do not occur in the corpus. This could be because they are either less known to the public and thus referred to in a different way or simply that due to the short length, it is less likely that something newsworthy has occurred there. Interesting is also that many of the locations that are not mentioned in the corpus are located in some of the newly developed areas such as Neu-Oerlikon or the Mühlacker in Affoltern. A possible reason for why they were never mentioned could be that some of the streets are relatively new or that they are referred to in a bigger context. For example, in the vicinity of the MFO-park are many (newer) street names are never mentioned in the corpus (e.g. Sophie-Taueber-Strasse, Meret-Oppenheimer-Weg, James-Joyce-Weg). This could be because the journalists talk about the bigger spatial context such as MFO-Park or Neu-Oerlikon instead of naming the streets, as they are more well known.

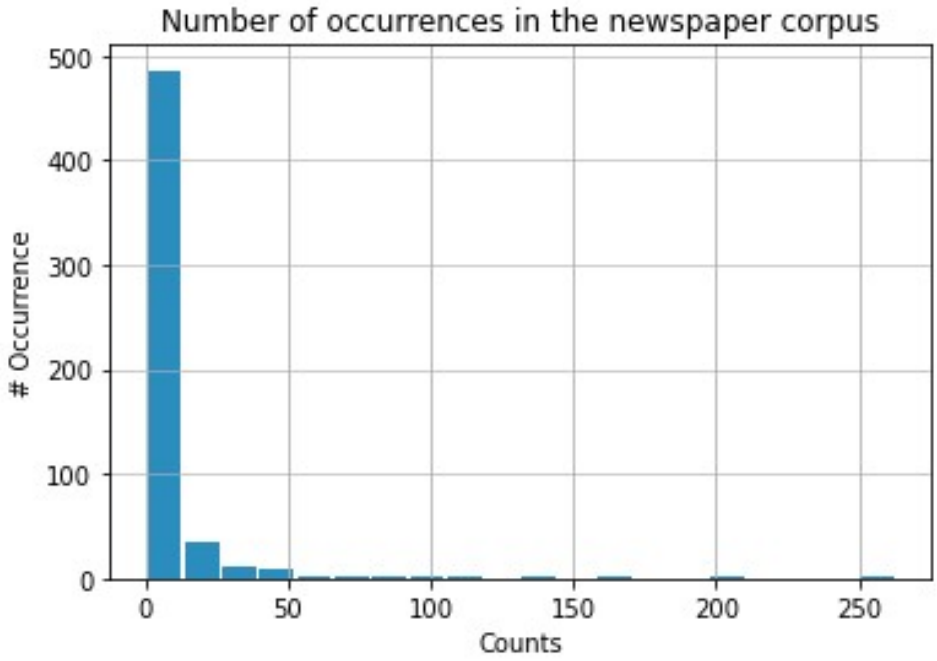

Figure 5: Histogram of the location frequencies in the newspaper corpus.

Looking at the distribution of location frequencies in the corpus, it is noticeable that most of the locations occur only very few times in the whole corpus, with the majority appearing less than ten times (Figure 5). Only 15 locations occur more than fifty times in the whole corpus when calculating the frequency with the TF-IDF method, with "Leutschenbach" having the highest occurrence with 208 references (Table 5). Interestingly, among these most mentioned locations some of the longest streets with the most inhabitants in Zurich can be found (Wehntalerstrasse: 5.7km, Winterthurerestrasse: 5.3km, Schafhauserstrasse: 4.8km) (Kohler *et al.*, 2016; Brusa, 2017). With the Leutschenbach, the Hagenholz and Thurgauerstrasse also three locations in the development area Leutschenbach are represented.

Table 5: Locations more than 50 times in the news corpus based on a TF-IDF weighted approach.

Location	Occurrence
Leutschenbach	208
Schaffhauserstrasse	166
Wehntalerstrasse	138
Winterthurerstrasse	111
Hagenholz	108
Einhausung Schwamendingen	93
Tramdepot	93
Thurgauerstrasse	87
Neu-Oerlikon	80
Birch	71
Wolfswinkel	68
Reckenholz	61
Schöneichtunnel	60
Milchbucktunnel	55
Mattenhof	51

4.1.2 Frequency of the Keywords over Time

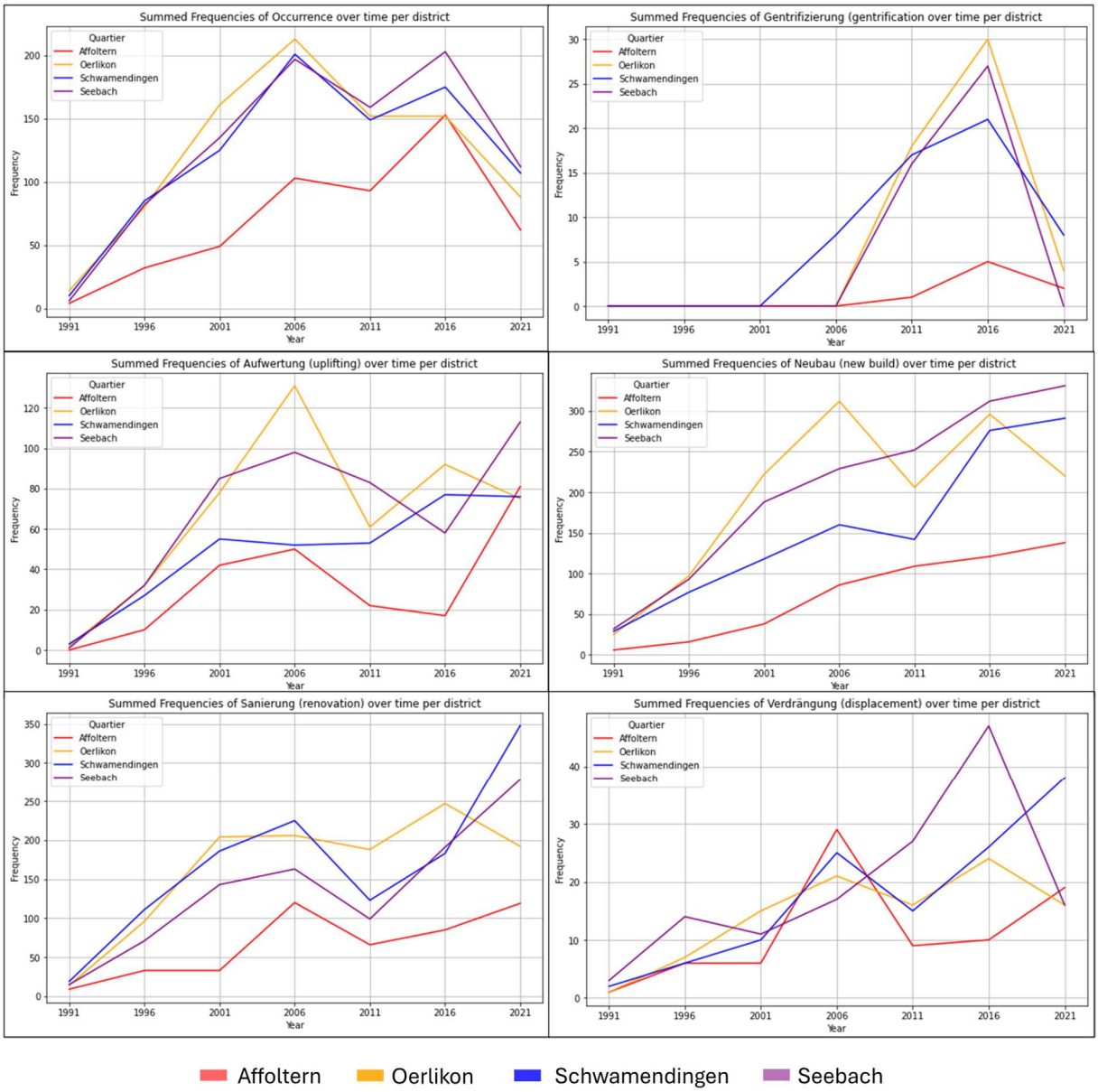

Figure 6: Summed frequencies of the occurrence and co-occurrence in the newspaper articles over time and per district.

The frequency of districts mentioned in the newspaper corpus and in relation to the keywords varies over time (Figure 6). Looking at the occurrence it is striking that Schwamendingen and Seebach have a very similar trajectory, with the highest peaks in 2006 and 2016, occurring up to 200 times. With over 200 occurrences, Oerlikon is mentioned the most in 2006 which corresponds to the development of Neu-Oerlikon. However, Oerlikon lacks the second peak in 2016 which mostly likely corresponds to the development of Leutschenbach which is located at the border of Schwamendingen and Seebach. Affoltern occurs the least over all time periods and has the highest peak in 2016 as well, with 150 occurrences.

Gentrifizierung (gentrification) is first used in the period of 2006 - 2010 in relation to Schwamendingen. The usage then steadily increases, although it remains quite low, peaking in the period 2016 - 2020 across all neighbourhoods, with a maximum frequency of 30 in Oerlikon. After 2021 the usage drops again, with Schwamendingen co-occurring the most often. Gentrification is used fewest in relation to locations in Affoltern (maximum of five occurrences) and highest with Oerlikon. The highest usage of *Aufwertung* (uplifting) is in Oerlikon in the period of 2006 – 2010 with a frequency of 130 which

corresponds to the development of Neu-Oerlikon. Afterwards, it drops again with a smaller peak in 2016. Interestingly, the trajectories of Seebach and Affoltern are almost identical although *Aufwertung* has generally been used more often in relation to Seebach (Seebach has a maximum frequency of 100 while Affoltern cooccurs a maximum of 50). In 2021, there is a sharp increase of usage in both these districts. In Schwamendingen, *Aufwertung* is used less often than for Oerlikon and Seebach, with the frequencies plateauing between 2001 and 2015 at 50 occurrences and between 2016 and 2024 at 80 co-occurrences. In Affoltern, Seebach and Schwamendingen the usage of *Neubau* (new build) steadily increases over time, reaching the peak co-occurrence in 2021 (Affoltern: 150, Schwamendingen: 300, Seebach: 330) reflecting the increased building activity in these districts. Oerlikon on the other hand has two distinctive peaks of 300 co-occurrences in 2006 and 2016. After 2021 the usage of *Neubau* decreases in Oerlikon perhaps indicating that less building activity is taking place there. Again, Affoltern has the lowest frequencies over all time periods. The occurrence of *Sanierung* (renovation) varies depending on the neighbourhood. Schwamendingen and Seebach have a very similar trajectory, with a steady increase until 2006 with occurrences of over 220 and 160 respectively. In 2016 the usage drops to just over 100 occurrences before rapidly increasing to 350 and 270 respectively in 2021. In Oerlikon *Sanierung* is used increasingly until 2001 where it levels at roughly 200 occurrences. Interestingly, the occurrence does not drop compared to the other districts, peaking again in 2016 at a frequency of 250 before dropping to just under 200 in 2021. Notable to other keywords, the peak in 2006 is missing. This reflects that Neu-Oerlikon was a previous industrial site where most of the buildings were demolished and built new rather than being renovated, thus explaining the missing peak. The usage of *Verdrängung* (displacement) is generally quite low. The first peak in usage occurs in 2006 in Affoltern, Schwamendingen and Oerlikon with frequencies between 20 to 30. The frequency of *Verdrängung* then drops to under 20 and in Affoltern even below 10, increasing again in 2016. While the frequencies of *Verdrängung* in Affoltern, Oerlikon and Schwamendingen are similar over time, Seebach differs. In Seebach, the occurrence of *Verdrängung* steadily increase until it peaks in the period 2016 – 2020 with a maximum occurrence of just over 40. This matches the construction of the Leutschenbach neighbourhood. After 2021 the highest frequency is found in Schwamendingen with an occurrence of just under 40. Analysing the total frequency over time and per district already gives a first valuable insight in the data, showing that the prevalence of the keywords varies not only over time but also neighbourhood. However, it is also heavily influenced by the total amount of articles per neighbourhood that were present in the corpus. To avoid this fallacy and gather understanding of how dominant urban transformation and gentrification is discussed in each neighbourhood, it is also important to look at the share of articles containing the keywords over time and in relation to the neighbourhoods (Figure 7).

Figure 7: Share of newspaper articles in the corpus that contain a certain keyword over time and per neighbourhood.

The graphs of the share of articles containing one of the keywords per neighbourhood give a different picture than the graph from above. To create the graphs, the occurrence data used above was divided by the total number of articles per period and neighbourhood. The graph of *Gentrifizierung* (gentrification) looks very similar to the total sum of usages distribution wise. Starting from 2006 onwards, the share peaks in 2016 where roughly 1% of all articles of Oerlikon, Schwamendingen and Seebach include gentrification. The highest share of articles containing gentrification is found in Oerlikon, then Schwamendingen, Seebach and Affoltern. In Affoltern, gentrification occurs only rarely, with a maximum occurrence in 0.2% of the articles in 2016. One of the most interesting districts in regard to *Aufwertung* (uplifting) is Affoltern, as it has the two highest peaks but also the lowest values of 1%. In 2021, almost 7% of the articles about Affoltern contain the word *Aufwertung*. The other peak is in 2001 where a bit less than 5% of all the articles contain *Aufwertung* while in 2016 less than 1% of the articles contain it. Seebach has a very similar trajectory than Afffoltern, but with less extremes. Compared to the two lows in 2016, Oerlikon and Schwamendingen have some of the highest values in this period with almost 4% of the articles using the word *Aufwertung*, perhaps reflecting the construction of the Mattenhof neighbourhood. The share of articles containing *Neubau* (new builid)

for Schwamendingen, Oerlikon and Seebach is always bigger than 6% with the lowest values in 1996. It then varies between 8% and 10% until 2015 and then rises up to more than 15% for Seebach and Schwamendingen in 2021. Interestingly, the large areal development of Neu-Oerlikon does not lead to a higher usage of *Neubau* in the corpus. While the other three districts have quite similar shares of coverage, the share of articles containing *Neubau* in Affoltern is below 6% before 2006 and only surpasses the 10% mark after 2021. Before that the highest peak is in 2011 with 8% which matches the construction activity at the Mühlackerstrasse. Remarkable is also that the share drops again to less than 6% in 2016 before increasing again. With a minimum of 4% of the articles and a maximum of over 18%, articles containing *Sanierung* (renovation) make up the highest share of articles across all periods. Interestingly, Schwamendingen has the highest share of occurrences across all periods, with a maximum of over 18% in 2021. An explanation could be the renovation of the Schöneichtunnel which took place during this time. Oerlikon, Schwamendingen and Seebach have again a very similar trajectory with the same peak in 2001 and low in 2011 (in Oerlikon the low is in 2006). The share in Affoltern is somewhat shifted with a peak in 2006 and low in 2016. Apart from Oerlikon, the share of articles containing *Sanierung* almost doubled between 2016 and 2021. Finally, only a small share of newspaper articles contain *Verdrängung* (displacement), with most of the periods having a share of roughly 0.75%. The most interesting are the peaks in 2006 in Affoltern, 2016 in Seebach and 2021 in Affoltern and Schwamendingen, where the share almost doubles to 1.5% - 2%. The higher share of articles after 2016 could imply the increased awareness around the issue of displacement in the media. Notable is also that the share of articles concerning *Verdrängung* in Oerlikon are always below 1%.

Figure 8: Thematic newspaper articles subdivided by keyword per neighbourhood from 1991 - 2021

When looking at the share of thematic newspaper articles in the corpus subdivided by neighbourhood (Figure 8), it becomes clear that the majority of articles in the corpus are not about gentrification or urban transformation. The share of article about urban development varies over time and by neighbourhood and ranges between 5% and 42%. Schwamendingen generally has the highest share of newspaper articles that are about urban transformation compared to the other neighbourhoods with the share always being above 15% and having the total maximum value of over 40% in 2021. With exception of 2021, Affoltern has the lowest share of newspaper coverage relating to urban

transformation usually around 10%. Two exceptions are 2006 (almost 20%) and 2021 (30%). With the exception of 1991 and 1996, the share of articles relating to the topic of urban transformation and gentrification in Oerlikon is quite stable over time ranging between 20% to just over 25%. The corpus about Seebach has a very similar share of thematic articles than Oerlikon except in 2021 where the share is much higher with almost 35% of all the articles relating to urban transformation and gentrification. When looking at the composition (Figure 8) of the thematic articles, most of them relate to either *Neubau* (new build) or *Sanierung* (renovation), with the majority ranging between 5% to 10%. Most interesting is the very high share of renovation articles in Schwamendingen in the period after 2021 where 19% of all the articles contain the word *Sanierung*. It can also be seen that *Gentrifizierung* (gentrification) and *Verdrängung* (displacement) only occur in very few articles in all the neighbourhoods. The highest value of just over 2% for displacement can be found in Schwamendingen in the newest period after 2021 while the highest values for gentrification can be found in Oerlikon and Schwamendingen in 2016. The share of articles containing *Aufwertung* (uplifting) is always below 5% with the exception of Affoltern (6.8%) and Seebach in 2021 (5.2%).

These simple graphs already show an interesting picture and indicate that the term usage depends not only on the period but also the location. They also show that more recently, in the period after 2021 the terms received more attention than in the previous years, showing their current relevance. This is why, in a next step, each of the different periods are analysed in more detail and compared to the actual building activity over time to get a deeper understanding of the temporal and spatial distribution of the newspaper coverage as well as howe well this relates to the construction activity.

4.1.3 Frequency and Co-Occurrence Results 1991 – 1995

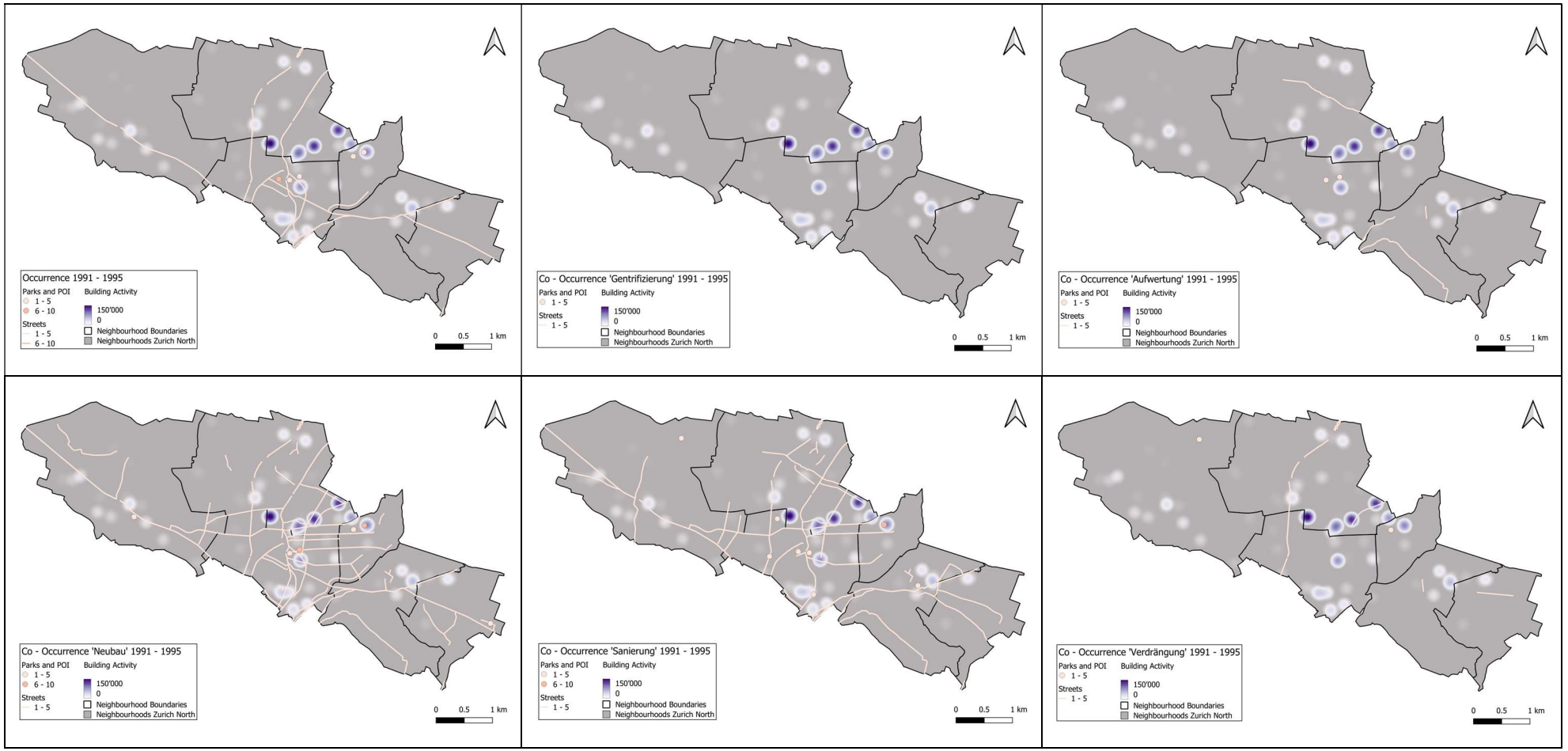

Figure 9: Frequency and co-occurrence in the newspaper corpus between 1991 – 1995 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 1991 - 1995

Between 1991 and 1995, there was only little building activity spread across all the three neighbourhoods (Figure 9). Standing out are the four distinct hotspots in the south of Seebach symbolising very high building activity. The hotspots represent mostly large office buildings such as the World Trade Centre at the Schärenmoos- and Leutschenbachstrasse or the university campus Zurich. Furthermore, construction was going on in the south of Oerlikon as well as around the Tramstrasse, the middle of Schwamendingen, the north of Seebach as well as along the Wehntalerstrasse in Affoltern. The effect of weighting the construction activity based on volume becomes visible in this example as single large-scale buildings are emphasised over smaller constructions.

Top 5 locations

In the top five locations for each of the maps are listed for a better comparison (Table 6). Remarkable are the very low frequencies of these top locations which is most likely connected to the small corpus containing only 937 articles with roughly 400'000 tokens during this period as well as the chosen TF-IDF method. The most frequent location mentioned in the period of 1991 – 1995 is the Marktplatz Oerlikon with 6 co-occurrences. This does not surprise in so far as it is the centre square in Oerlikon with many shops. Interesting is also that 3 out of the top five most frequently occurring locations are some of the longest streets in Zurich which highlights their importance in the region. The Leutschenbach, which ranks third in this period, appears often due to articles about TV-events and shows, since the national radio and television studios are located there. In the period 1991 – 1995 gentrification does not occur in the corpus thus there are also no co-occurrence results. For *Aufwertung* (uplifting) it is remarkable that the maximum frequency is only one, indicating that it is not a word used often during that time. Only seven articles include the word in this period and only two of them relate to the uplifting of an urban area however not in Zurich North. *Neubau* (new build) on the other hand co-occurs more often with locations in Zurich North, most of all with Messgelände, Hagenholz and Leutschenbach respectively the streets they are located at. These places are mentioned in the corpus mostly in regard to the planned construction of the event hall in Oerlikon in the second half of the 1990s, the opening of the World Trade Centre in Leutschenbach as well as the extension of the incinerator facility Hagenholz. Interestingly, these locations are all in the vicinity of each other and are often mentioned in the same article. *Sanierung* (renovation) is mentioned most often with the Hagenholz, again in relation to work on the incinerator facility. The other locations are mentioned in the corpus in regard to credits granted by the city council to renovate different buildings and public infrastructure. Notable is that the project located on the Zürichbergstrasse is not located in Zurich North but more towards the city. This is an issue with longer streets that pass through different districts in general and that is unaccounted for in with my approach. Lastly, *Verdrängung* (displacement) which also only has very low frequencies co-occurs the most with Leutschenbach in this period in relation again to the TV-studio. However, displacement does not appear in the context of displacement of persons, which is of interest in this work, but in the context of one subject being displaced by another. Looking whether some locations co-occur with multiple of the keywords only a few places stand out: namely the Messegelände, Hagenholz and Leutschenbach.

Table 6: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 1991 - 1995.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Marktplatz Oerlikon (6)		Messegelände (1)	Messegelände (9)	Hagenholz (8)	Leutschenbach (2)
2	Wehntaler-strasse (4)		Am Katzenbach (1)	Hagenholz (7)	Birchstrasse (3)	Reckenholz (1)
3	Leutschen-bach (3)		Sternen Oerlikon (1)	Wallisellen-strasse (5)	Zürichbergstrasse (3)	Friedrichstrasse (1)
4	Schaffhauser-strasse (3)		Frohburgstrasse (1)	Hagenholz-strasse (5)	Zehntenhaus-strasse (3)	Birchstrasse (1)
5	Winterthurer-strasse (3)		Friedrichstrasse (1)	Leutschenbach (4)	Binzmühlestrasse (2)	Leutschenbach-strasse (1)

Comparison between the newspaper coverage and the construction activity

The frequencies of locations in the newspaper does not correlate statistically significantly with the building activity between 1991 – 1995 statistically, except in Affoltern (Table 7). Furthermore, the analysis shows that neither the usage of *Aufwertung* (uplifting) nor *Verdrängung* (displacement) correlates with the building activity which supports the results of the close reading that these two words are not used in relation to urban transformation. The correlation between *Neubau* (new build) and the construction activity on the other hand is statistically significant in a weak positive manner in all data subsections. This makes sense in so far as *Neubau* directly refers to construction however, it also indicates that the data in the corpus actually refers to locations in Zurich North and not to other places. *Sanierung* (renovation) also has a positive correlation with the building activity in some places. Interestingly, the locations in Seebach even correlate moderately positive with the building activity during this time which could be explained by Hagenholz which was both extended (new built) as well as renovated during this time. Co-occurrences in Affoltern and Schwamendingen do not correlate significantly with the building activity. Since the GWR does not include information about renovations and since renovations can also be related to infrastructure projects such as water pipes or electricity lines, this is explicable. Despite the statistically significant correlation between *Neubau* and the construction activity, it is interesting to note that none of the streets where the construction hotspots are located at received any media attention despite the Hagenholz. A reason for this could be that while the buildings have a high volume, they are single buildings developed by private developers that are neither discussed in city council meetings nor have a large impact on a neighbourhood. Furthermore, the articles about the World Trade Centre tend to use the name of the building instead of the street it is located at. Since I did not include the World Trade Centre as a location in the gazetteer, potential co-occurrences would have been missed.

Table 7: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity in the period 1991 - 1995. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	0.04		-0.03	0.17	0.14	0.01
Non-profit	0.12		-0.06	0.18	0.2	-0.01
For-profit	-0.07		-0.02	0.13	0.1	0.01
Affoltern	0.17			0.12	0.18	-0.05
Oerlikon	0.09		-0.07	0.27	0.16	0.09
Schwamendingen	0.1		0.02	0.18	0.07	0.06
Seebach	0		-0.05	0.26	0.3	0

4.1.4 Frequency and Co-Occurrence Results 1996 – 2000

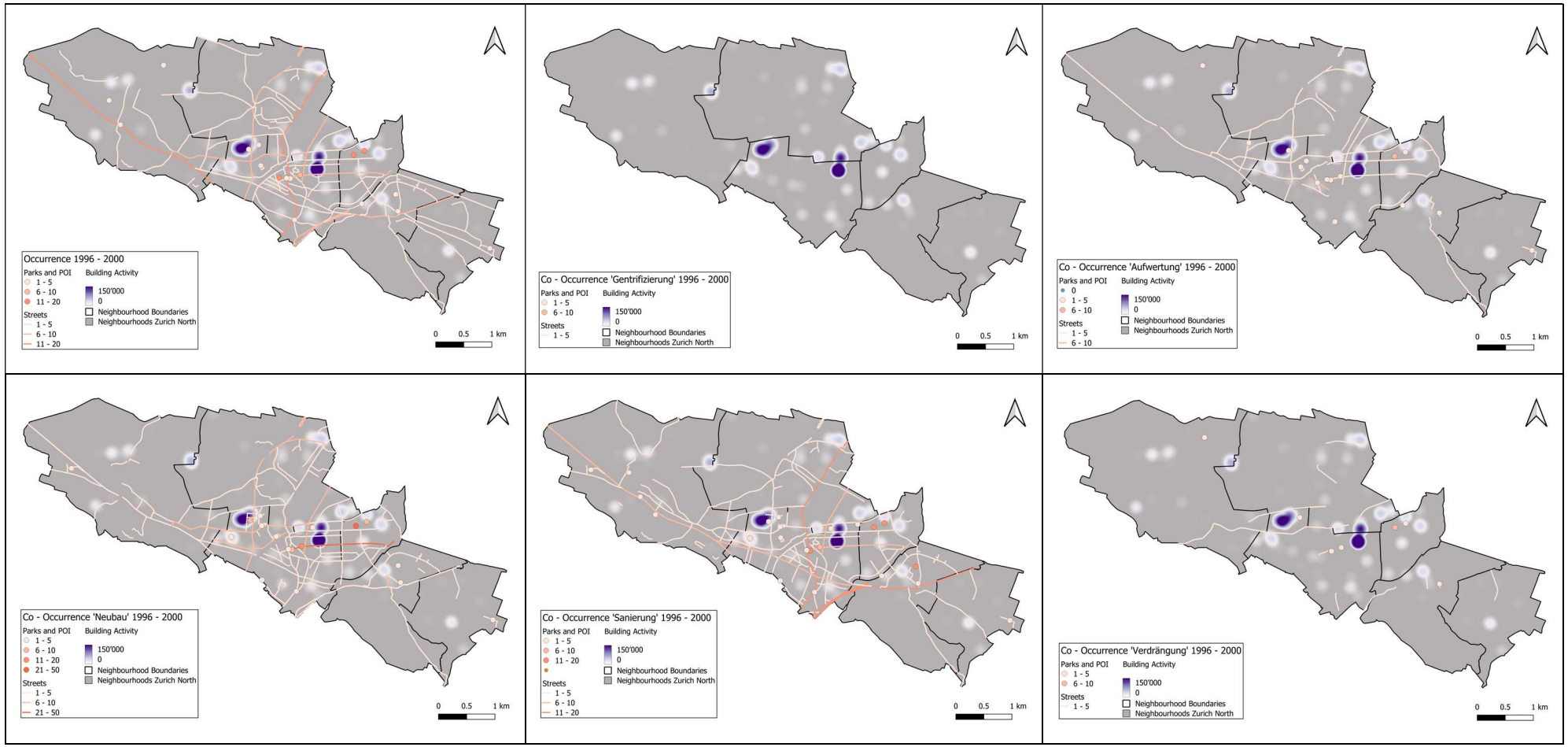

Figure 10: Frequency and co-occurrence in the newspaper corpus between 1996 – 2000 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 1996 - 2000

In the period 1996 – 2000 there are three hotspots regarding building activity of which all of them are located in Oerlikon (Figure 10). In the northwest of Oerlikon at the Eduard-Imhof-Strasse a big warehouse was built in 1997 and in the northeast at the Wallisellenstrasse the Messegelände Zürich (an event hall) in combination with a big car park was constructed in 1998. Besides these hotspots, there was only little building activity occurring for example in the northeast and southeast of Seebach. In Affoltern and Schwamendingen very little buildings were constructed during this period.

Top 5 locations

The top five frequencies and co-occurrences for each category are generally higher than in the period before, most likely reflecting the bigger corpus that was available at that time (Table 8). The most occurring locations in the corpus are very similar than in the previous period, again reflecting their importance in the region apart from Hagenholz that oust the Winterthurerstrasse to the sixth place. The same as the period before, there are no results for Gentrifizierung (gentrification). *Aufwertung* (uplifting) is most used with locations in the Leutschenbach area (Leutschenbach and Thurgauerstrasse), around the Marktplatz Oerlikon (Franklinstrasse), before the Messegelände (Wallisellenstrasse) and at the Schwamendingerplatz (Table 8). However, none of these places co-occur more than ten times, reflecting that the word is only used in 72 articles during that period. The most frequent co-occurrences for *Neubau* (new build) are again the Leutschenbach, where many articles talk about it as a possible location for the new football stadium, as well as the Messegelände, Wallisellenstrasse and Hagenholz, where newspaper articles talk about the incinerator facility in general and whether or not it must be extended again. *Sanierung* (renovation) also co-occurs the most with Hagenholz and Leutschenbach, again talking about investment in the incinerator facility and the new football stadium. Interesting is the co-occurrence of Schöneichtunnel which is covered in the newspaper articles in two regards: on one hand the pending and urgent renovation of the tunnel in 2001 which cannot be postponed and will lead to blockages on the other hand about an initiative that sought to enclose the national highway A1 with a so-called Einhausung. Since the Einhausung ended up being built and will be finished in 2025, I will also look more closely at the usages of Einhausung in the collocation analysis. However, it should be kept in mind that already in 1996, the idea and term existed. *Verdrängung* (displacement) has similarly to *Aufwertung* again very low frequencies although still higher than in in the period before. Again, the Leutschenbach and Messegelände are found in the top five besides locations that have not co-occurred with other keywords before. At the Reckenholz, the national research institute for agriculture Agroscope is located. The reason why it co-occurs so often is that news articles report on the loss of biodiversity and the “displacement” of native plants and animals due to the intense land use and invasive species. Also, at the Schwamendingerplatz and the Binzmühlestrasse *Verdrängung* is not used in relation to people but to small businesses being ousted by a major distributor and traffic. Looking again at the co-occurrence between a location and multiple keywords, it is again Leutschenbach, Hagenholz and the Messegelände / Wallisellenstrasse that stand out.

Table 8: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 1996 - 2000.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Marktplatz Oerlikon (16)		Leutschenbach (9)	Leutschen-bach (29)	Hagenholz (17)	Leutschenbach (9)
2	Leutschen-bach (16)		Franklinstrasse (6)	Wallisellen-strasse (21)	Leutschen-bach (14)	Reckenholz (5)
3	Schaffhauser-strasse (15)		Wallisellen-strasse (5)	Messegelände (18)	Schöneich-tunnel (14)	Messegelände (3)
4	Wehntaler-strasse (12)		Thurgauerstrasse (5)	Hagenholz (10)	Tramdepot (13)	Schwamendinger-platz (2)
5	Hagenholz (11)		Schwamendinger-platz (5)	Tramdepot (9)	Schaffhauser-strasse (13)	Binzmühle-strasse (1)

Comparison between the newspaper coverage and the construction activity

Comparing the newspaper coverage and the construction activity shows that they correlate weak positively but statistically significant with Neubau in the whole corpus as well as the non—profit corpus (Table 9). Furthermore, there is a statistical significance between the construction activity in Seebach and the general occurrence of locations in the corpus. Striking is also that the construction of the Messegelände was reported on by the media, shown by the high occurrence of both Messegelände and Wallisellenstrasse. In contrast to this match, the other big construction during this period at the Eduard-Imhof-Strasse was not reported on at all. Perhaps the new construction did not grab media attention since it is an industrial warehouse built in an industrial area that only has started transforming into Neu-Oerlikon. Another mismatch between the newspaper coverage and the building activity is Leutschenbach on which I now want to have a closer look by close reading some of the articles about it.

Leutschenbach is in the top five frequencies of all the categories. Leutschenbach is a located in the south of Seebach bordering both Oerlikon and Schwamendingen. Reading some articles that were published during this time, reveals that there are both public and private plans to uplift and develop the Leutschenbach region where planners see enormous potential due to the largeness of the area and the vicinity to the Oerlikon train station and airport. Already in 1996, a group of private investors and landowners joined forces to develop a plan for the Leutschenbach region as they criticised that the city only focused on developing the industrial site Zentrum Zürich Nord (Neu-Oerlikon) and Zurich West despite "Leutschenbach's obvious qualities" (NZZ., 1996). By 1998, newspapers cover the topic of development plans and concepts that are being created for industrial brownfields in the city of Zurich where Leutschenbach is also officially mentioned. However, the focus during this time lay more on the Zentrum Zürich Nord where the planning process was already further developed. Interestingly, after covering the brownfield development plans in a more neutral and informative way, already in February 2000 the potential impact of the new neighbourhoods on Alt-Oerlikon (the old centre) were discussed. The journalist summarised:

> "*In die Euphorie über das Zentrum Zürich Nord, die Stadtplaner und Investoren erfasst hat, mögen nicht alle alteingesessenen Oerliker einstimmen. Die Entstehung eines neuen Stadtquartiers wird mit einer Mischung von Wohlwollen und Skepsis verfolgt.* " (NZZ., 2000)
>
> (In the euphoria about the Zurich North centre that has gripped urban planners and investors, not all long-established residents of Oerlikon are in favour. The emergence of a new urban neighbourhood is being followed with a mixture of benevolence and scepticism.)

Afterwards, in December of the same year the city of Zurich formally published a development plan for Leutschenbach, laying out the future of the neighbourhood (City of Zurich, 2000). However, this was not really discussed in the newspaper also in the following year. This example shows that the process to develop a new neighbourhood is long and starts long before the first buildings are completed. Nonetheless, newspaper already report on some of these first steps in the long process

Table 9: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity in the period 1996 - 2000. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	0.00		0.05	0.08	0.00	-0.01
Non-profit	-0.04		0.09	0.18	0.05	0.00
For-profit	0.03		0.04	0.05	-0.03	-0.01
Affoltern	0.05		0.12	0.15	0.06	0.07
Oerlikon	-0.09		0.04	0.08	-0.10	0.00
Schwamendingen	-0.06		0.07	0.02	-0.02	-0.01
Seebach	0.21		0.11	0.09	0.06	0.13

4.1.5 Frequency and Co-Occurrence Results 2001 – 2005

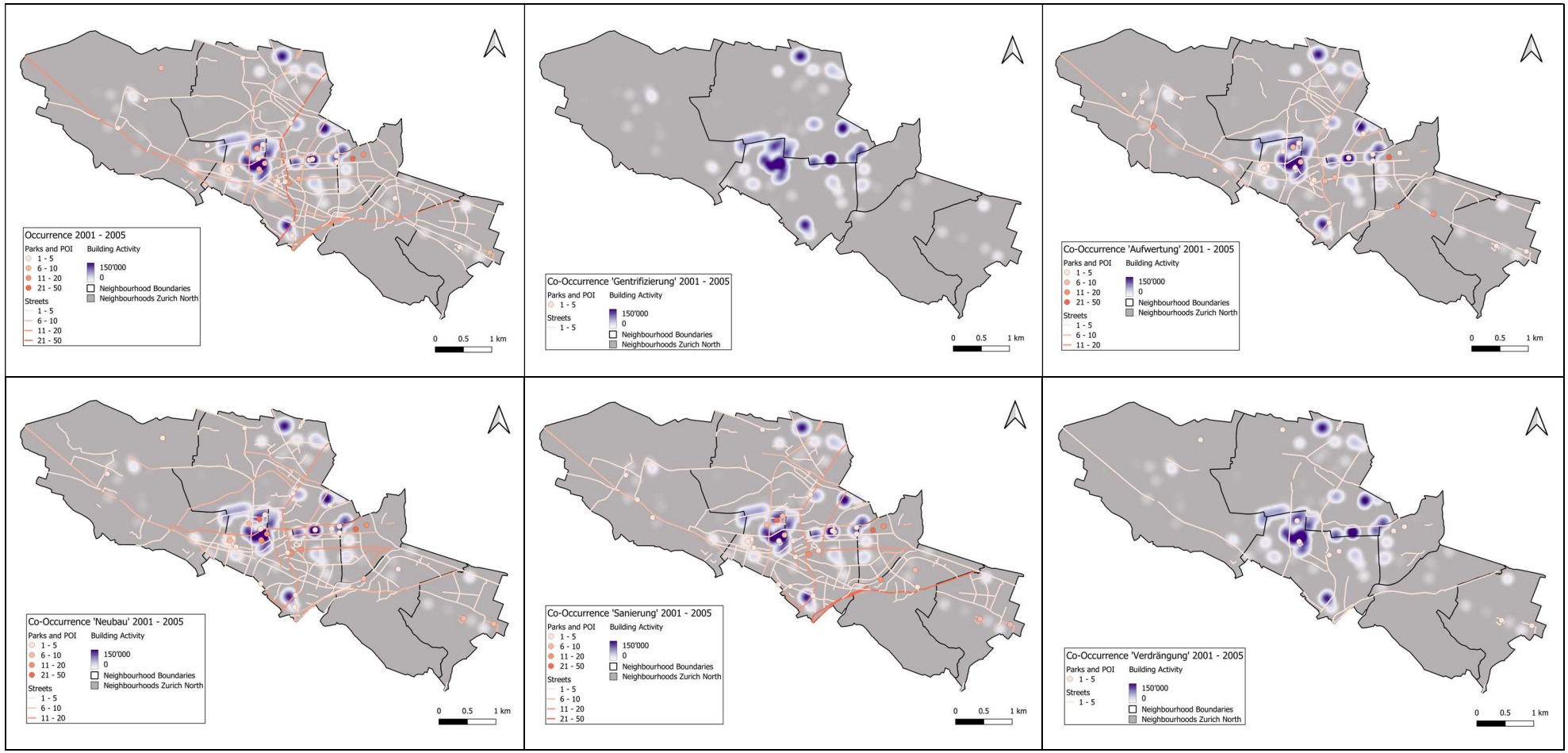

Figure 11: Frequency and co-occurrence in the newspaper corpus between 2001 – 2005 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 2001 – 2005

At the beginning of the new millennium, construction activity in Zurich North takes off. Between 2001 – 2005, the biggest construction activity hotspot is located in the north of Oerlikon right next to the train station and opposite the old centre (Figure 11), where the new district Neu-Oerlikon is created on the 60-hectare big industrial brownfield of the former machine factory Bührle. The areal transformation is one of the biggest urban development areas in Switzerland up to date (City of Zurich, 2024f). Another big building activity hotspot in Oerlikon is in the south at the Schaffhauserstrasse, where the City Bernina is located. The City Bernina opened in 2001 and is a mixed-use complex that contains apartments, businesses and offices. Besides Oerlikon, big construction hotspots are also located in Seebach. At the border to Oerlikon and Schwamendingen at the Hagenholzstrasse, the construction of two high-rise buildings was finished at the Quadro Platz. With 88 and 72 meters hight respectively, the two towers are some of the highest buildings in Zurich, housing offices for large cooperations (Architonic, 2024). This newbuild was the first step in the development of the Leutschenbach neighbourhood is supposed to be the entrance to the new district and a symbol for the coming change (City of Zurich, 2024g). Also, in the Hagenholzstrasse, around the Hunziker-Areal and Andreaspark, the first construction projects began, and first offices and housing blocks were ready for occupancy. Additionally, construction projects were finished in the north of Seebach where the Eichrain residential area was built as well as the west where a big office building was constructed. Neither in Affoltern nor in Schwamendingen large building activity hotspots exist.

Top 5 locations

The top five locations that were mentioned most frequently in the corpus are very similar than in the previous periods and include again Leutschenbach as well as the three long streets (Table 10). Appearing for the first time in the top five is Birch which is located in the development area Neu-Oerlikon. Birch is often mentioned in regard the development plans in Neu-Oerlikon in general and the school called im Birch that was built during this time more specifically. Birch is also mentioned often due to its vicinity of one of the four parks that were created within the Neu-Oerlikon development: Wahlenpark, Oerliker-Park, Louis-Häfliger Park and MFO-Park. As in the previous years, there are no results for *Gentrifizierung* (gentrification). *Aufwertung* (displacement) again co-occurs the most often with Leutschenbach which can be explained again by the development plan for Leutschenbach. Interestingly, with Zehntenhausplatz and Schwamendingerplatz the two centres of Affoltern and Schwamendingen are mentioned despite that there is not a lot of building activity. Upon close reading it becomes clear that the city of Zurich published a new transportation plan that seeks to encourage people to walk, by amongst other things calming traffic and uplifting multiple local neighbourhood centres. Furthermore, newspapers report on the planned development plans in Affoltern, which are at an early stage at this time. The plan is to develop multiple various green- and brownfields such as the Ruggächer, in Büngert and CeCe-areal into residential areas. Intriguing is also the mentioning of the Einhausung Schwamendingen. This is because the initiative to cover the highway that is dividing the neighbourhood is picking up speed and more support in the government. Filtering articles that contain both Einhausung and *Aufwertung* interesting articles can be found that talk about the bad public image, problems, low rents and the low social status of Schwamendingen and efforts of private groups and the city to uplift the neighbourhood and image:

> "*Die Stadt will mithelfen, das Quartier positiv weiterzuentwickeln. Wenn beispielsweise in der geplanten Kulturbeiz etwas Attraktives zu Stande kommt, geht man eben am Abend nach Schwamendingen - so wie heute ins Lettenquartier. Warum soll nicht auch in Schwamendingen plötzlich etwas "Schickes" entstehen? Es ist durchaus möglich, dass Junge auf der Suche nach neuen Wohnformen sagen, jetzt gehen wir extra nach Schwamendingen wohnen.*" (Eitle, 2001)

> (The city wants to help develop the neighbourhood in a positive way. If, for example, something attractive comes about in the planned cultural centre, people will simply go to Schwamendingen in the evening - just like they do today in the Letten district. Why shouldn't something “chic” suddenly emerge in Schwamendingen too? It is quite possible that young people in search of new forms of living will say, now we're going to live in Schwamendingen.)

Compared to the other three neighbourhoods, no concrete development plan yet exists for the transformation but only guiding principles for the development. It is interesting, that at this point articles write about the need for uplifting in the neighbourhood which is seen as something positive.

Neubau (new build) co-occurs the most with Birch and Neu-Oerlikon as the papers report on the creation of the new neighbourhood in general and single projects such as the school more specifically. Again, Leutschenbach is co-occurring with *Neubau* a lot due to the development plans and first completed stages. The high frequencies of both the Schulstrasse and Tramdepot are misleading as most articles do not talk about the locations in Oerlikon but locations in other places with the same name. This shows the problem of unambiguous location names. The results for *Sanierung* (renovation) are very similar than in the previous years while reflecting the development of Neu-Oerlikon. Again, the Schöneichtunnel co-occurs often due to its renovation at that time. *Verdrängung* (displacement) has again very low frequencies with the Reckenholz occurring most often and articles talking about different topics, the same as in the period before.

Table 10: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 2001 - 2005.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Leutschen-bach (27)		Leutschenbach (30)	Birch (50)	Birch (44)	Reckenholz (3)
2	Schaffhauser-strasse (21)		Zehntenhaus-platz (15)	Leutschen-bach (41)	Leutschen-bach (35)	Messegelände (3)
3	Birch (17)		Einhausung (12)	Neu-Oerlikon (19)	Schöneich-tunnel (34)	Leutschen-bach (2)
4	Wehntaler-strasse (15)		Schwamendinger-platz (11)	Schulstrasse (17)	Winterthurer-strasse (24)	Birch (2)
5	Winterthurer-strasse (15)		Milchbucktunnel (9)	Tramdepot (15)	Hagenholz (20)	Hagenholz-strasse (2)

Comparison between the newspaper coverage and the construction activity

In the period 2001 – 2005, the newspaper coverage in Zurich North does not correlate with the construction activity apart from the data subsection of locations in Affoltern (Table 11). This is somewhat surprising since there was only very little construction activity and occurring at that time. Looking at Figure 11 though, it can be seen that in most of the places where construction was going on, there was some newspaper coverage even though perhaps less than expected around the development of Neu-Oerlikon, the Quadro-Platz and City Bernina. As an example of a location where there is a lot of construction activity but no coverage at all, it is noteworthy that except for the Binzmühlestrasse, none of the streets located close to the centre of Neu-Oerlikon and the MFO-Park are mentioned in the corpus during this time. Possibly this can be explained by the fact that most of these streets were either not built yet or still very new. Therefore, newspaper articles most likely refer to the area within the bigger context of Neu-Oerlikon or the most prominent POIs in the area such as Birch or MFO-Park. Oppositely, the location Sunnige Hof in the east of Schwamendingen occur frequently in the corpus during the period despite little to no building activity in the area. After further

investigation, it turns out that both of the terms are not unambiguous. Sunnige Hof for example is not only the name of a housing cooperative estate in Schwamendingen but also the name of the housing cooperative itself. Although the cooperative is most active in Schwamendingen it also owns estates in other districts which could cause the gap between coverage and occurrence.

Table 11: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity in the period 2001 - 2005. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	0.04		0.04	0.07	0.05	0.01
Non-profit	0.13		0.12	0.14	0.10	-0.07
For-profit	0.00		0.00	0.05	0.03	0.05
Affoltern	0.29		0.06	0.20	0.16	0.21
Oerlikon	-0.07		-0.01	0.00	-0.06	-0.03
Schwamendingen	0.07		0.06	0.05	0.05	0.01
Seebach	0.07		0.04	0.04	0.12	0.03

4.1.6 Frequency and Co-Occurrence Results 2006 – 2010

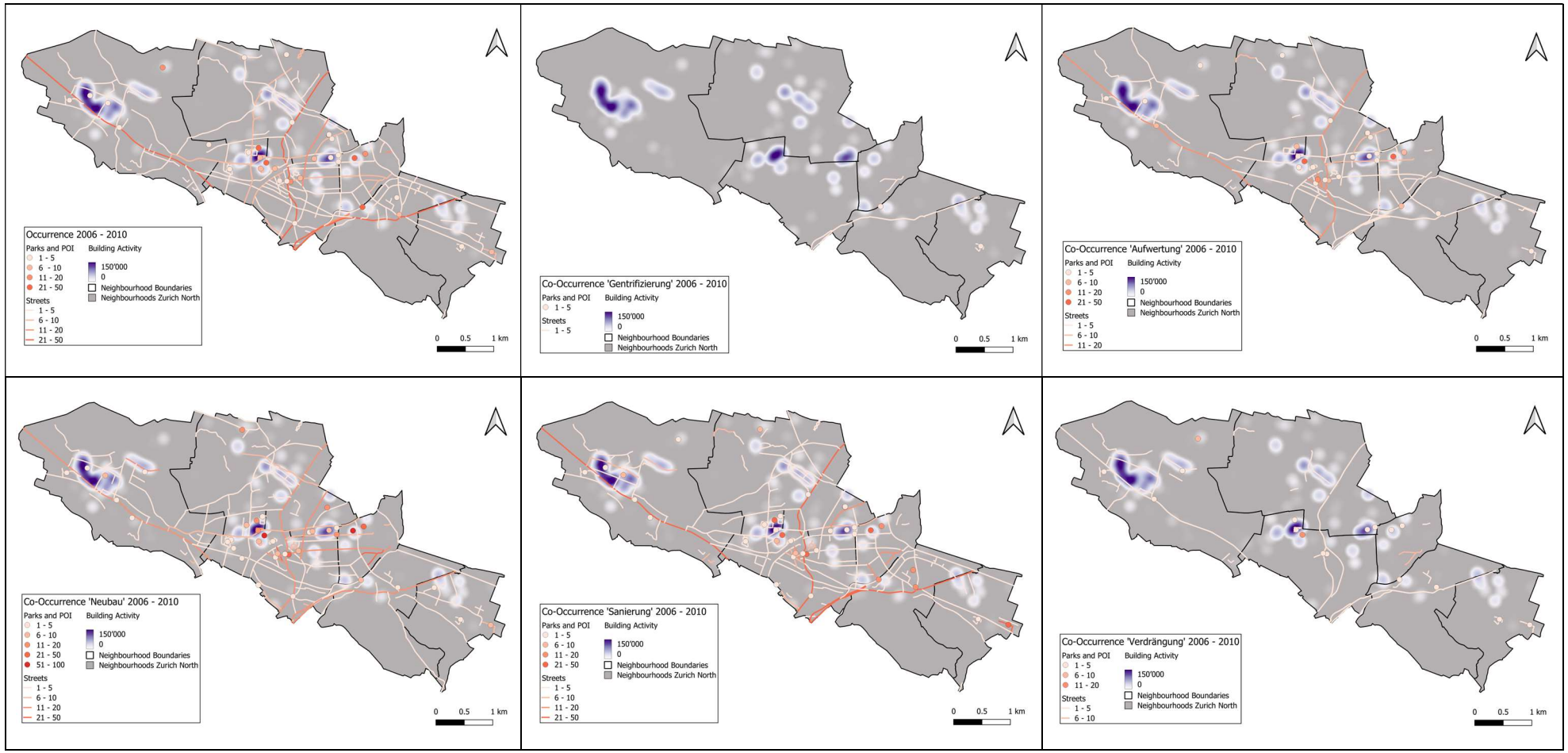

Figure 12: Frequency and co-occurrence in the newspaper corpus between 2006 – 2010 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 2006 – 2010

The most construction activity between 2006 – 2010 is occurring in the north-west of Affoltern, where multiple large Überbauungen (areal developments) were built on some of the last large land reserves in Zurich (ABZ, 2024): *In Büngerten*, *Ruggächern*, *Wolfswinkel*, *Aspholz*, *Klee* and the former industrial site of a graphite factory at the *Wehntalerstrasse*. Between 2006 and 2010, 5'000 people moved into these new residential areas at the edge of Affoltern, increasing the population by almost 25% from 18'800 to almost 23'000 (NZZ, 2008). A lot of these projects were undertaken by housing cooperatives on former industrial sites and have been complimented on the design. Other large development areas are again in the region Neu-Oerlikon especially around the MFO park. Also, in the Leutschenbach, Hagenholz and Hunziker-Areal area, more construction is going on as the new district is being built. In the middle of Seebach the cooperative housing *am Katzenbach 1 & 2* was finished as replacement construction the old multi-family homes from the 1940s (City of Zurich, 2024e). In Schwamendingen there are also a few construction projects, namely at the Luegislandstrasse and the Funkwiesenstrasse that were both conducted by a housing cooperative (Figure 12).

Top 5 locations

In the period 2006 – 2010 the most frequent locations in the corpus are again very similar than in the previous periods except that Neu-Oerlikon now also makes it to the top five, indicating that the neighbourhood and its name are becoming more established. For the first time, *Gentrifizierung* (gentrification) is used in the corpus in connection with the Sunnige Hof, the Winterthurerstrasse, the Einhausung and the Dreispitz (Table 12). Interestingly, they are all located in Schwamendingen; however, they do not match at all with the construction activity occurring at that time. Taking a closer look at the newspaper corpus brings more clarity to the somewhat surprising result: In the period between 2006 – 2010 there are in total only two articles that contain *Gentrifizierung* (gentrification). One is about a Swiss arts museum in New York, the other is a 2010 article from the NZZ about the image problem of Schwamendingen (Troxler, 2010). It is the latter article that led to the co-occurrence results containing the sentence *"Von der in der Innenstadt kritisierten Gentrifizierung findet man hier keine Spur"* (There is no trace here [in Schwamendingen] of the gentrification criticised in the city centre). Instead, the article reports about the problems of housing cooperatives to find a good mixture of inhabitants for their buildings, since only few Swiss families apply for their apartments and the proportion of foreign nationals is high. Furthermore, the article says that despite the dry housing market, many people living in the inner city would not consider moving to Schwamendingen due to the bad image of the district, a feared ghettoisation and a decreasing social status. The article focuses on different housing cooperative projects (Sunnige Hof, Winterthurerstrasse, Dreispitz) that are "turning things around" for Schwamendingen, as the new constructions manage to attract a good earning "urban audience" and "well integrated foreign nationals" especially due to housing shortage in other districts. The Einhausung is mentioned as pushing the upturn forward as many housing cooperatives plan new constructions and renovations when the project is finalised, improving the quality of the housing to more modern standards. Interestingly, the problem of *Verdrängung* (displacement) is also addressed by one city representative, stating:

> *"Schwamendingen [hat] die Voraussetzungen für eine gute Entwicklung […], ohne dass die unteren Einkommensschichten ganz verdrängt werden."* (Troxler, 2010)
>
> (Schwamendingen has the prerequisites for good development without lower income groups being completely displaced)

This example shows the caveat of the co-occurrence methods since it does not account for negations. However, even though the article does not talk about gentrification in Schwamendingen in general, it

is still an interesting source as it describes the process of uplifting especially with the housing cooperative projects as desired and welcomed and puts it in contrast to the "current ghettoisation" of the district. *Aufwertung* (uplifting) is co-occurring the most with Neu-Oerlikon, Leutschenbach as well as places around the Marktplatz in Oerlikon which was uplifted during this time with new illumination, benches and a fountain. With over 60 co-occurrences, Neu-Oerlikon and Leutschenbach occur the most with *Neubau* (new build). This is not surprising since the two neighbourhoods are developed during this time. Other places that co-occur often are the Tramdepot, Hagenholz and Dreispitz. *Sanierung* (renovation) co-occurs most often with the Wehntalerstrasse which could be due to the developments in Affoltern. Interestingly, Leutschenbach, Neu-Oerlikon and Birch also appear in the top five even though buildings are built new rather than being renovated most likely due to simultaneous infrastructure renovation. Lastly, *Verdrängung* (displacement) co-occurred the most with Neu-Oerlikon in this time period. Looking at some of the articles, displacement is again used for a variety of topics. However, for the first time a few articles also talk about how urbanisation and densification can lead to the displacement of people although Neu-Oerlikon is only mentioned as a side note (one article is about Zug another about the housing market and loos off affordable apartments in Zurich in general). Notable, displacement in these articles is always brought up by academic research in either sociology or urban geography.

Table 12: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 2006 - 2010.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Leutschen-bach (44)	Einhausung (2)	Neu-Oerlikon (35)	Neu-Oerlikon (78)	Wehntaler-strasse (35)	Neu-Oerlikon (11)
2	Schaffhauser-strasse (38)	Dreispitz (2)	Leutschen-bach (29)	Leutschenbach (61)	Winterthurer-strasse (31)	Reckenholz (8)
3	Neu-Oerlikon (33)	Sunnige Hof (2)	Marktplatz Oerlikon (14)	Tramdepot (33)	Leutschen-bach (30)	Dreispitz (6)
4	Wehntaler-strasse (26)	Eigengrund (2)	Querstrasse (13)	Hagenholz (27)	Neu-Oerlikon (30)	Felsberg (5)
5	Winterthurer-strasse (25)	Winterthurer-strasse (2)	Schaffhauser-strasse (12)	Dreispitz (21)	Birch (25)	Leutschenbach (3)

Comparison between the newspaper coverage and the construction activity

Generally, when looking at the maps (Figure 12) it seems that where a lot of building activity is occurring, newspapers report on it using *Neubau*, *Sanierung* and *Aufwertung.* However, statistically no significant correlation can be determined in either of the different data subsections (Table 13). The places that receive the most attention are the newly built areas Neu-Oerlikon, with the point of interests MFO, Oerliker Park, Birch and Max-Bill-Platz as well as the development area of Leutschenbach with its corresponding POI Hagenholz, Andreaspark, Mehr als Wohnen and Leutschenbachpark. More interesting however, the large areal development in the west of Affoltern receives only little attention despite its size, its similarity to Neu-Oerlikon and Leutschenbach development and its impact on the population growth in Affoltern. A possible explanation could be that Neu-Oerlikon and Leutschenbach are located more centrally and thus receive more attention. As in the previous periods, longer streets that also lead to other districts such as the Winterthurerstrasse or Schaffhauserstrasse as well as centrally located places such as the Marktplatz Oerlikon also co-occur often, even though there are no building projects occurring in the vicinity of the street, indicating their general importance in the region.

Table 13: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity in the period 2006 - 2010. In this period none of the correlations between the coverage and the building activity are statistically significant (p< 0.05).

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	-0.01	-0.04	-0.04	0.04	0.02	-0.04
Non-profit	0.06	-0.06	-0.02	0.02	0.03	-0.12
For-profit	-0.06	-0.04	-0.05	0.03	0.01	-0.01
Affoltern	0.03		0.05	0.07	0.13	-0.04
Oerlikon	-0.04	0.04	-0.12	-0.01	-0.06	-0.05
Schwamendingen	0.08	-0.08	0.05	0.11	0.08	-0.07
Seebach	0.03		0.02	0.10	0.10	0.00

4.1.7 Frequency and Co-Occurrence Results 2011 – 2015

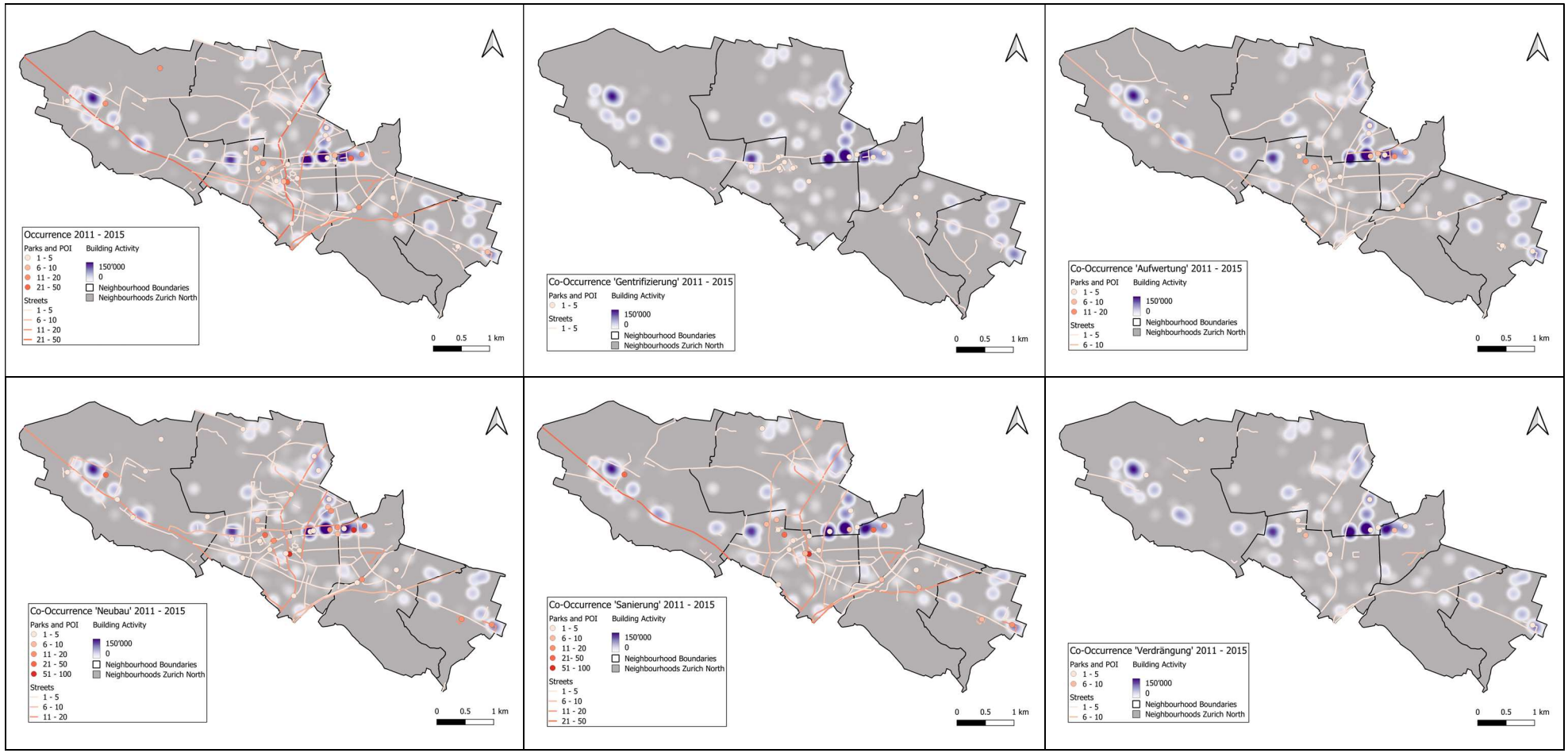

Figure 13: Frequency and co-occurrence in the newspaper corpus between 2011 – 2015 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 2011 – 2015

In the period 2011 – 2015, the most construction activity took place at the border between Seebach and Schwamendingen where the Leutschenbach neighbourhood developed (Figure 13). Leutschenbach was built on a former industrial site on plans developed between 2000 and 2012. During this period both residential housing (Mehr als Wohnen) as well as office towers in the area of Quadro Platz (Sunrise Tower and Zurich Insurance Company) were built (City of Zurich, 2024c). In Affoltern, the biggest construction activity took place in the west at the Mühlackerstrasse (the housing cooperative estate Siedlung Klee VIII), where already many buildings were finished in the previous period. Another housing cooperation project in Affoltern during this time was the Furttalstrasse, the In Böden Siedlung and the Mötteliweg Siedlung located at the Binzmühlestrasse. In Oerlikon, the highest building activity volume wise was the "creative business centre" NOERD in Neu-Oerlikon that houses different creative companies from marketing, design, software engineering and TV production (NOERD, 2024) and another industrial building in the area. In Seebach multiple housing cooperative projects were finalised such as the Köschenrüti in the North, the extension of the Katzenbach in the centre of Seebach, the Birchstrasse in the north as well as the Schaffhauserstrasse in the east. In Schwamendingen the biggest construction activity took place in the east where part of the Mattenhof estate belonging to the housing cooperative Sunnige Hof was finalised. Furthermore, housing cooperatives carried out replacement projects at the Luegislandstrasse where already in the previous years, replacement buildings were finalised.

Top 5 locations

In the period 2011 – 2014 the locations with the highest frequencies are Leutschenbach and Hagenholz, the Wehntaler- and Schaffhauserstrasse as well as the Tramdepot (Table 14). This is not surprising given that the Leutschenbach / Hagenholz were built during this period. While still occurring often in the corpus, places in Neu-Oerlikon do not make it to the top five, indicating that the development is finished there and the neighbourhood established. *Gentrifizierung* (Gentrification) occurs more often in this period than ever before, namely for 30 different places, although none of the locations co-occurs more than five times. This could imply that the term gentrification has slowly found its way to Zurich and Zurich North. The highest co-occurrences are with locations in Neu-Oerlikon and Leutschenbach. Especially interesting is that Mehr als Wohnen appears in the top five given that it is a housing cooperative that is non-profit driven. In total there are 29 articles that contain gentrification in the corpus. Most of these articles talk about gentrification in connection with the Europaallee and the Kalkbreite and not directly with Neu-Oerlikon or the Leutschenbach area. However, gentrification is related to large areal developments all around Switzerland in general and Neu-Oerlikon is often mentioned as an example in an article the journalist states:

> *"Was die einen Gentrifizierung nennen, nennt die SBB Immobilien AG "urbanen Genuss."* (Keller and Bodmer, 2013)
>
> (What some call gentrification, the SBB real estate company call "urban indulgence.)

Other articles, report on creatives locating to Neu-Oerlikon because they are displaced from the Kreis 4 and because the area is still edgy. In the article the journalist comments that this development as *"Verheissung oder aber als Vorboten einer drohenden Gentrifizierung"* (promise or as a harbinger of impending gentrification). (Kucera, 2011). Leutschenbach is brought up in relation to an identified linguistic gentrification of the city as a bus station is renamed from Kehrichtverbrennungsanlage (incinerator facility) to Genossenschaftsstrasse (cooperative street) (Rohrer, 2015). Two of the articles about gentrification relate to Schwamendingen and portray a local artist there who states that the project of the Einhausung that will cause a lot of displacement and will uplift the neighbourhood. The

article states that with these developments, gentrification has set in (Gimes, 2015). Close reading yet again shows that while the chosen method is helpful to spot places of change and general developments it can also be misleading as it neglects the context, nuances, negations and cross references. *Aufwertung* (uplifting) co-occurs the most again with the two big development areas Leutschenbach and Neu-Oerlikon. Interestingly, two housing cooperative projects can also be found in the top five co-occurrences with uplifting. Again, the Einhausung is mentioned, showing the importance of the project for Schwamendingen. The top five for *Neubau* (new build) and *Sanierung* (renovation) are again very similar to each other and also to the previous years, showing the connectedness of the two terms. With the Blumenfeld, another school building project made it in the top five frequencies, showing that they these types of projects cause a lot of news coverage. *Verdrängung* (displacement) co-occurs often with three large areal development: Neu-Oerlikon, Leutschenbach and Dreispitz (in Basel) but is again not only used in connection to people but also businesses, green spaces and other topics. Interestingly, Mehr als Wohnen also co-occurs as in some articles it is mentioned as a positive example to counter the displacement of lower income groups of people since the housing cooperative also provides large apartments for reasonably low prices. However, the results must be treated with caution since the frequencies are low and sometimes caused by a duplication of articles.

Table 14: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 2011 - 2015.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Leutschen-bach (43)	Leutschenbach (4)	Leutschenbach (20)	Leutschen-bach (75)	Tramdepot (54)	Neu-Oerlikon (9)
2	Wehntaler-strasse (31)	Neu-Oerlikon (4)	Neu-Oerlikon (12)	Tramdepot (63)	Leutschen-bach (30)	Leutschen-bach (8)
3	Tramdepot (27)	Mehr als Wohnen (4)	Mehr als Wohnen (10)	Neu-Oerlikon (47)	Wehntaler-strasse (26)	Dreispitz (7)
4	Schaffhauser-strasse (27)	Genossenschafts-strasse (4)	Kirchenfeld (10)	Blumenfeld (47)	Neu-Oerlikon (25)	Rundweg (4)
5	Hagenholz (20)	Binzmühle-strasse (3)	Einhausung (9)	Hagenholz (22)	Blumenfeld (22)	Mehr als Wohnen (3)

Comparison between the newspaper coverage and the construction activity

In the period 2011 – 2015 the newspaper coverage correlates with the building activity in a statistically significant weak positive way for some of the data subsets (Table 15). Most strikingly the coverage about locations in Seebach correlates significantly for all keywords except Verdrängung (displacement). This most likely relates to the development of the Leutschenbach area which caught a lot of media attention. Furthermore, the correlation over all locations is significant for the frequency, Aufwertung (uplifting), Neubau (new build) and Sanierung (renovation). Despite of a few large, finalised housing cooperative projects, there was again not a lot of newspaper coverage in Affoltern and the correlation with the building activity not significant. Also, in Oerlikon the correlation is not significant which could be because of the relatively high media coverage on Neu-Oerlikon even after the construction is finished, implying that such large projects remain relevant even after the creation. In Schwamendingen the correlation between the co-occurrence of Aufwertung (uplifting) and Neubau (newbuild) and the construction activity is statistically significant most likely also due to the coverage on the Leutschenbach development. Astonishing is the mismatch between the coverage of Dreispitz in relation to all keywords except gentrification despite the lacking building activity. After a short investigation it becomes apparent that in the news, they do not talk about the Dreispitz in

Schwamendingen but the large areal development in Basel. The articles were extracted because they contained both “Zürich” and “Dreispitz” since Zurich is used as reference point for other large urban development project.

Table 15: Results of the Spearman’s' Rho analysis of the correlation between newspaper coverage and building activity in the period 2011 – 2015. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	0.13	0.02	0.10	0.13	0.09	0.03
Non-profit	0.12	-0.06	0.13	0.09	0.09	0.08
For-profit	0.10	0.04	0.08	0.13	0.06	-0.01
Affoltern	0.12	-0.04	0.04	0.13	0.05	0.04
Oerlikon	0.08	-0.07	0.06	0.03	0.04	-0.10
Schwamendingen	0.17	0.01	0.17	0.19	0.08	-0.04
Seebach	0.23	0.17	0.25	0.30	0.19	0.15

4.1.8 Frequency and Co-Occurrence Results 2016 – 2020

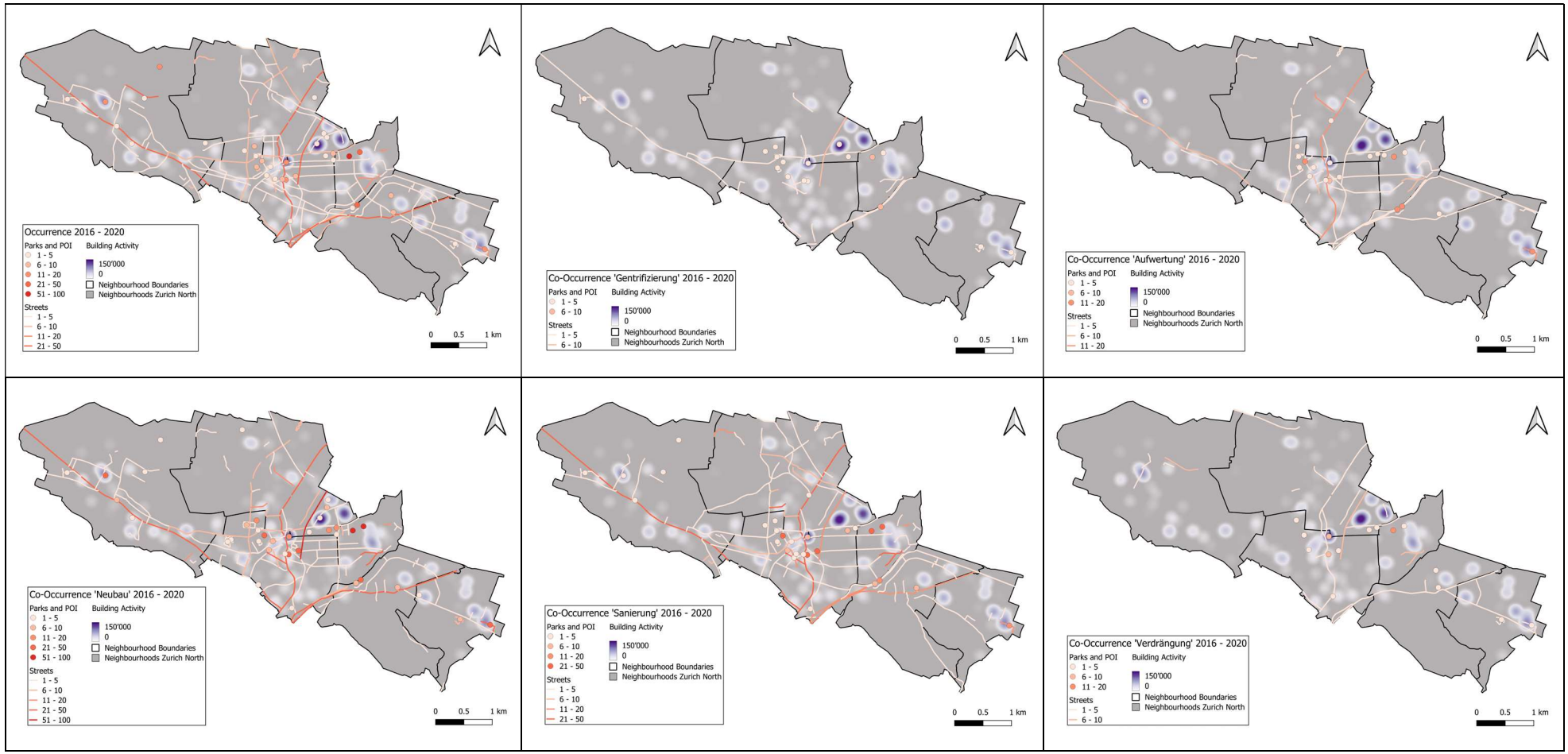

Figure 14: Frequency and co-occurrence in the newspaper corpus between 2016 – 2020 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 2016 – 2020

Between 2016 and 2020 there was a lot of building activity in all four districts although most of the big areal developments such as the Neu-Oerlikon, Leutschenbach or Mühlacker were completed (Figure 14). Instead, private landowners and smaller housing cooperatives started to realise projects. In Schwamendingen the biggest construction activity was occurring at Mattenhof in the east and Am Glattbogen in the north, both by housing cooperatives. In Seebach the main building activity was again in the Leutschenbach area, where the residential tower Wolkenwerk was constructed. In Oerlikon there is not one large hotspot but large spread smaller private building activity, with the biggest project being the 80m high Andreasturm, an office tower right next to the train station. In Affoltern, further residential houses were developed in the area of Mühlacker in the west by private landowners as well as along the Wehntalerstrasse.

Top 5 locations

Compared to the previous periods, Wolfswinkel and Thurgauerstrasse appear in in the top five frequencies in the period 2016 – 2020 besides Leutschenbach and Wehntalerstrasse which generally occur often in the corpus (Table 16). Out of the 97 articles in total that contain the word *Gentrifizierung* (gentrification), the majority (54) were published during this period with 35 places co-occuring in total. The places that co-occur the most are the Einhausung and the Thurgauerstrasse which are both mentioned 7 times each. This can be explained by the construction start for the Einhausung in 2019 and a referendum vote on part of the layout plan at the Thurgauerstrasse. The referendum was taken as the plan to build multiple high-rise buildings up to 70m was criticised for being too high, the public participation too low and the project in general not enough socially responsible. A journalist comments:

> "*Bei der Linken paart sich die Kritik an der baulichen Verdichtung jeweils sofort mit dem Abwehrreflex gegenüber jeder Art von Gentrifizierung […] Und in Zürich Nord hat eine ungewöhnliche Allianz […] ein neues Stadtquartier an der Thurgauerstrasse ausgebremst*.» (Troxler, 2019)
>
> (On the left, criticism of building densification is immediately coupled with a defence reflex against any kind of gentrification […] And in Zurich North, an unusual alliance […] has put the brakes on a new urban neighbourhood on Thurgauerstrasse)

However, in most articles, gentrification again is directly related to the Langstrasse or the Europallee and the Thurgauerstrasse is mentioned merely as a side note. In contrast to that, the issue of gentrification is actively discussed in relation to the Einhausung. Thereby, various stakeholders such as the city council are quoted, most of them seeing no danger of gentrification, as the housing co-operatives ensure apartments with low rents. While they recognise that the project will ultimately lead to rent increases which will have an impact on the demographic, this is seen as a positive and desirable outcome: "*Eine bessere Durchmischung ist gut für Schwamendingen*" (Fritzsche, 2018) (a better social mixing is good for Schwamendingen). *Aufwertung* (uplifting) co-occurs most often with the Einhausung and the Überlandpark which will be created on top of it which shows the positive effect the project is expected to have on the neighbourhood. Also, in the top five co-occurrences Mattenhof can be found which matches the building activity in that region. *Neubau* (new build) again co-occurs the most often with Leutschenbach, Hagenholz and Thurgauerstrasse which are mentioned in relation to the public votes, the still ongoing development in the neighbourhood as well as the school Leutschenbach which is often brought up in articles about other new school projects in the city. The Messegelände also appears in the top five, however all the articles are about the event hall in St. Gallen, showing again the issue of ambiguous location names. The top five results for *Sanierung* (renovation) are very similar than in the previous periods, containing long streets and large development areas where various

renovation works is conducted. While still only having low co-occurrences, *Verdrängung* (displacement) appears most often in this period, namely with the Leutschenbach and the Thurgauerstrasse. However, the problem of the variety of different kinds of displacement remains. For example, the term is used in connection to politicians, businesses, and problems. Nonetheless, a few articles can be found that talk about densification, gentrification and displacement as an issue. However, they do so in a broad way, reporting on the issue in the city in general, as well as measures against it and not in regard to specific locations.

Table 16: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 2016 - 2020.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Leutschenbach (54)	Einhausung (7)	Einhausung (18)	Leutschenbach (93)	Leutschen-bach (42)	Leutschen-bach (15)
2	Wolfswinkel (46)	Thurgauer-strasse (7)	Schaffhauser-strasse (14)	Hagenholz (70)	Wehntaler-strasse (41)	Thurgauer-strasse (10)
3	Wehntaler-strasse (33)	Leutschen-bach (6)	Mattenhof (14)	Thurgauer-strasse (51)	Tramdepot (39)	Tramdepot (10)
4	Schaffhauser-strasse (30)	Neu-Oerlikon (4)	Überlandpark (14)	Messegelände (43)	Dreispitz (37)	Dreispitz (9)
5	Thurgauer-strasse (30)	Mehr als Wohnen (4)	Leutschenbach (13)	Tramdepot (39)	Schaffhauser-strasse (34)	Wolfswinkel (7)

Comparison between the newspaper coverage and the construction activity

Generally, there is no statistically significant correlation between the newspaper coverage and the construction activity in the period 2016 – 2020 apart from the locations in Affoltern that correlate weak positively with *Aufwertung* (uplifting) and *Neubau* (new build) and Seebach that correlates weak positively with *Gentrifizierung* (gentrification) which most likely relates to the coverage and activity in the Leutschenbach area (Table 17). Despite the lacking statistical correlation many of the construction sites and co-occurrences match very well such as the Blumenfeld, Mattenhof, Andreasturm, Wolkenwerk or Kirchenfeld that received at least some attention by the media. Looking at locations where the coverage and the building activity do not match, one of the most striking places is the Wolfswinkel which also appeared in the top five frequencies and *Verdrängung* (displacement) co-occurrences. After a short investigation, however, it becomes clear that in the media, the Wolfswinkel does not relate to the area in Affoltern but to a football player called Wolfswinkel. Conversely, the housing cooperative Am Glattbogen received no media coverage despite having similar dimensions to the Mattenhof and being built at the same time.

Table 17: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity in the period 2016 - 2020. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	0.05	0.02	0.04	0.02	0.03	0.04
Non-profit	0.06	0.08	0.11	0.01	0.01	0.04
For-profit	0.00	-0.01	-0.01	-0.01	0.01	0.02
Affoltern	0.16	0.09	0.20	0.20	0.09	0.01
Oerlikon	0.10	0.00	0.06	0.00	0.07	0.14
Schwamendingen	-0.05	-0.10	-0.11	-0.10	-0.07	-0.10
Seebach	0.13	0.16	0.10	0.10	0.11	0.09

4.1.9 Frequency and Co-Occurrence Results 2021 – 2024

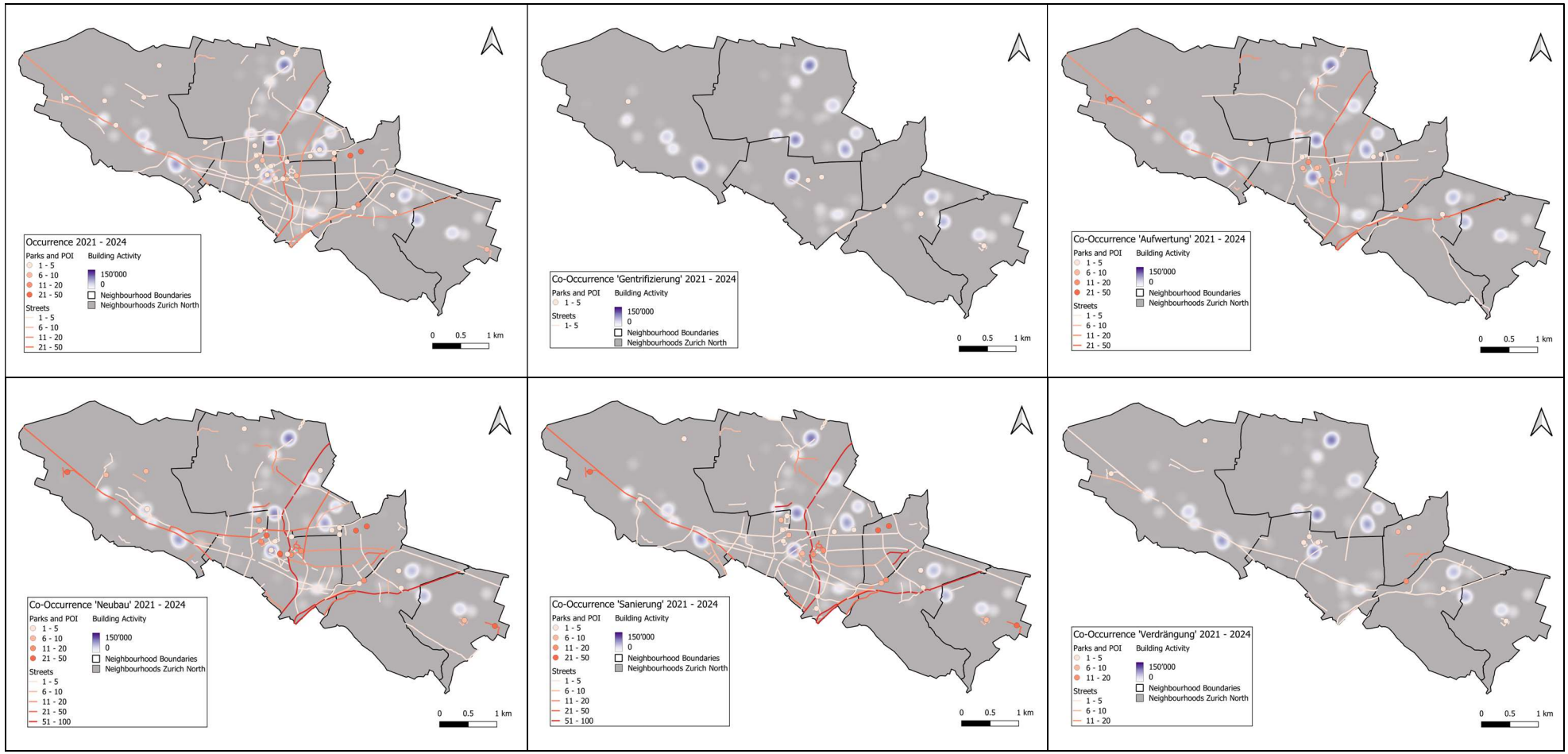

Figure 15: Frequency and co-occurrence in the newspaper corpus between 2021 – 2024 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 2021 – 2024

In the period between 2021 and 2024 there is less building activity than in the previous years and there are only a few hotspots in each of the four neighbourhoods (Figure 15). In Affoltern, the biggest building activity takes place in the southeast at the Wehntaler- / Obsthaldenstrasse as well again in the In Böden housing cooperative. In the north of Seebach another areal development, the Eichrain residence was built by the city of Zurich at the Birchstrasse / Glatttalstrasse in combination with a residence and health centre for the elderly and a kindergarten (City of Zurich, 2024d). In the south of Seebach the areal development at the Eggbühlstrasse was built by private owners. During this time some more high-rise towers were finished, namely the Tower 55 in Leutschenbach at the Hagenholzstrasse as well as the Franklinturm in Oerlikon which is described as "*Zusammen mit dem Andreasturm wird er die Skyline des boomenden Stadtteils prägen*" (together with the Andreasturm, it [Franklinturm] will characterise the skyline of the booming district) (Franklinturm, 2024). In Schwamendingen multiple housing cooperatives finished replacement constructions for example a the Glattstegweg and Riedacker.

Top 5 locations

The top five most frequent locations present in the corpus between 2021 – 2024 is very similar to the previous year with the exception that the Einhausung is in the top five (Table 18). This reflects again the ongoing development in Leutschenbach, the final stage of the Einhausung project, which will be finished by 2025 and the general relevance of the Schaffhauser- and Winterthurerstrasse. During this period there are 18 articles that contain *Gentrifizierung* (gentrification). However, upon close reading it becomes clear that many of the articles either talk about the phenomenon in general or about the gentrification of the Langstrasse and Europaallee. This shows that despite the large transformation the neighbourhoods undergo, gentrification is not a word often used in the context of Zurich North. A few articles, however, ask the question of whether the project will set a gentrification wave in process. Thereby, they interview different people such as the president of the neighbourhood association or the chair of the district 12 interest group. While both acknowledge that the rents will rise and that some long-time residents will be displaced due to the neighbourhood uplifting caused by the Einhausung, they both state that they don't believe that gentrification will set in due to the large share of housing cooperatives that will continue to provide apartments for low-income households. In both articles they express their belief that the uplifting of Schwamendingen is a positive development and that they welcome it (Fritzsche, 2018; SRF, 2023). *Aufwertung* (uplifting) co-occurs most often with the Winterthurer- and Schaffhauserstrasse due to diverse uplifting measures along these roads (both building and nature wise). Besides the Einhausung and the Thurgauerstrasse, Holzerhurd appears in the top five co-occurrences. This is explicable considering the discussions around the newly planned tramway line in Affoltern. *Neubau* (new build) again co-occurs most often with the Schaffhauser- and Winterthurerstrasse, again in relation to various smaller projects along these roads. Also, the Leutschenbach and the Hagenholz co-occur frequently as the neighbourhood is being developed further. The high co-occurrence value for Frohburgstrasse is surprising, but it turns out that it is caused by a peculiarity in the dataset. Thereby, short notifications in news ticker formats that are published within a week are not removed but stacked on top of each other and published again under a new headline, date and ID. This leads to a multiplication of certain articles which leads to high frequencies in the co-occurrence analysis. These types of articles are very hard to filter since neither the date, the title, the ID nor the first characters are identical. The *Sanierung* (renovation) results contain again the Schaffhauser- and the Winterthurerstrasse indicating that renovations, newbuilds and uplifting measures are often interconnected and conducted at the same time. Additionally, the Rickenstrasse in Seebach and the Mattenhof and Dreispitz in Schwamendingen co-occur very frequently also due to a multiplication of articles. *Verdrängung* (displacement) appears in a total of 84 articles whereby most of them are again not about the displacement of inhabitants but other topics. In Leutschenbach for

example, the articles report on the emergence of invasive species that "displace" native animals, or in connection with the TV-studio while Dreispitz refers again to the areal development in Basel. In Neu-Oerlikon, *Verdrängung* relates mostly to trade businesses that, according to some, were not considered enough during the establishment of the neighbourhood. Only the articles containing Einhausung use displacement in connection with inhabitants but do not identify it as the sole cause but as a part of the uplifting process in Schwamendingen. The Furttalstrasse has such a high co-occurrence value also due to the multiplication of ticker-news.

Table 18: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 2021 - 2024.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Leutschenbach (23)	Einhausung (3)	Winterthurer-strasse (31)	Schaffhauser-strasse (74)	Schaffhauser-strasse (95)	Dreispitz (12)
2	Schaffhauser-strasse (23)	Schöneich-tunnel (2)	Schaffhauser-strasse (30)	Winterthurer-strasse (65)	Winterthurer-strasse (70)	Einhausung (11)
3	Hagenholz (23)	Sunnige Hof (2)	Holzerhurd (21)	Leutschen-bach (49)	Rickenstrasse (59)	Leutschen-bach (10)
4	Winterthurer-strasse (19)	Messegelände (2)	Einhausung (16)	Hagenholz (45)	Dreispitz (57)	Furttalstrasse (9)
5	Einhausung (17)	Blumenfeld (2)	Thurgauer-strasse (16)	Frohburg-strasse (44)	Mattenhof (50)	Neu-Oerlikon (5)

Comparison between the newspaper coverage and the construction activity

In the period 2021 – 2024 there is no statistically significant correlation between the newspaper coverage and the construction activity for any of the data subsets (Table 19). This supports the above finding where many of the articles are not related to the building activity in that area. When comparing the news coverage and building activity visually (Figure 15) the only areas that do match are along the Wehntalerstrasse and the middle of Seebach. Looking at places where the coverage does not match the building activity, the Einhausung is one of the most prominent places. However, the discrepancy is explicable since the GWR only contains data about buildings and not infrastructure projects. Therefore, the Einhausung would not show up in the GWR. Examples for a location with high building activity but no coverage are the housing cooperative estates Obsthalde and In Böden which did not receive any media attention despite creating roughly 200 more apartments in a time where more living room is needed (SGE, 2024; Waidmatt, 2024). Interestingly, the Franklinturm received similar attention in this period than before even though the project was completed in 2022. This implies that reporting on construction projects is often at its highest during the planning and authorisation process. Once the construction has been approved, the public discourse subsides and with it the media coverage.

Table 19: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity in the period 2021 - 2024. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	-0.03	-0.04	-0.04	0.00	0.02	0.02
Non-profit	-0.09	-0.08	-0.03	-0.03	0.00	-0.03
For-profit	-0.02	-0.03	-0.05	-0.01	0.01	0.03
Affoltern	0.01	-0.06	-0.01	0.13	0.11	0.02
Oerlikon	-0.03	0.04	-0.01	0.00	0.02	0.08
Schwamendingen	-0.05	-0.10	-0.11	-0.10	-0.07	-0.10
Seebach	0.09		0.08	0.07	0.11	0.13

4.1.10 Frequency and Co-Occurrence Results 1991 – 2024

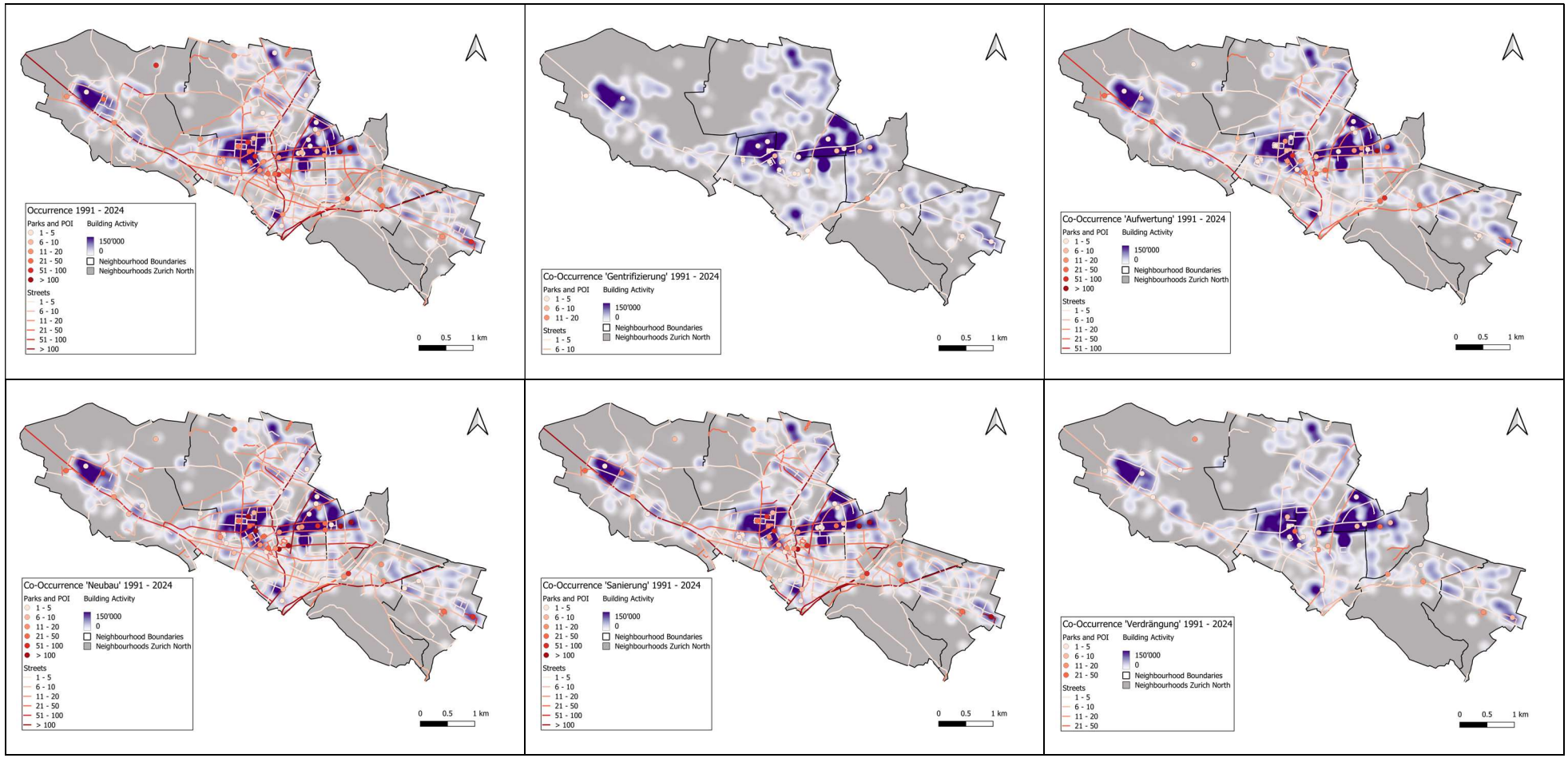

Figure 16: Frequency and co-occurrence in the newspaper corpus between 1991 – 2024 for the five keywords Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) compared to the building activity weighted by volume during this time. The frequency and co-occurrence results are displayed in orange (the redder, the more frequent the location occurs in the corpus) and the building activity in purple (the more violet, the higher the building activity).

Construction activity 1991 – 2024

Over the past 30-years, Zurich North experienced major constructional transformations (Figure 16). Most prominently three large industrial brown sites were developed, creating the new neighbourhoods Neu-Oerlikon, Mühlacker and Leutschenbach. Besides that, and not less important, many replacement construction projects were conducted exhaustively across all neighbourhoods, changing the appearance and character of many neighbourhoods.

Top 5 locations

Not surprisingly, the top five locations over the whole time period are very similar than in the individual periods and can be divided into four categories: Firstly, the Leutschenbach and surrounding area (Hagenholz, Thurgauerstrasse and Mehr als Wohnen) that has the highest occurrence in the corpus for four out of the six periods analysed. This can be explained on one hand by the large areal development, the creation of the neighbourhood as well as the construction of the school led to a high coverage in the news. On the other hand, Leutschenbach is where the national television studios are located at, which is also often mentioned in the news. The second category concerns the areal development of Neu-Oerlikon which also ranks high in all the categories except for the frequencies. This shows that these large developments are the dominant talking points of media about Zurich North. The third category contains the three long streets Schaffhauserstrasse, Wehntalerstrase and Winterthurerstrasse. Because of their length and relevance to the neighbourhoods, they appear frequently in the corpus in general and especially regarding infrastructure and building projects. The last category concerns the Einhausung which co-occurs the most with gentrification and uplifting. As the project was initiated in the 1990s it is mentioned in all periods and more recently in sense of the impact it will have on Schwamendingen. In the top 5 over all periods also the Tramdepot, Reckenholz and Dreispitz are mentioned however, they have nothing to do with urban development in Zurich North (Table 20).

Table 20: Top five locations for each of the analysis dimensions: frequency, Gentrifizierung (gentrification), Aufwertung (uplifting), Neubau (new build), Sanierung (renovation) and Verdrängung (displacement) in the period between 1991 - 2024.

	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
1	Leutschenbach (208)	Einhausung (14)	Leutschenbach (109)	Leutschen-bach (352)	Schaffhauser-strasse (192)	Leutschen-bach (49)
2	Schaffhauser-strasse (166)	Leutschen-bach (10)	Neu-Oerlikon (80)	Neu-Oerlikon (206)	Leutschenbach (178)	Dreispitz (37)
3	Wehntaler-strasse (138)	Neu-Oerlikon (8)	Schaffhauser-strasse (69)	Hagenholz (192)	Winterthurer-strasse (172)	Neu-Oerlikon (26)
4	Winterthurer-strasse (111)	Mehr als Wohnen (8)	Einhausung (66)	Tramdepot (178)	Tramdepot (162)	Reckenholz (20)
5	Hagenholz (108)	Thurgauer-strasse (7)	Wehntaler-strasse (53)	Schaffhauser-strasse (170)	Hagenholz (145)	Thurgauer-strasse (15)

Comparison between the newspaper coverage and the construction activity

Over the entire period the newspaper coverage correlates statistically significantly and weakly positively with the construction activity for some of the analysed categories and locations (Table 21). Most strikingly, the news coverage and building activity is statistically significant in all the categories for the for-profit locations and locations in Seebach. The statistical significance for locations in Seebach is most likely due to the high news coverage and building activity in the Leutschenbach area. Interestingly, however, a significant correlation is not given for locations in Oerlikon or Affoltern, despite the Neu-Oerlikon and Mühlacker areal developments being quite similar. The statistical significance of

the for-profit locations could also be linked to the construction of Leutschenbach, where many for-profit buildings were built. Overall, the newspaper coverage in general as well as about *Aufwertung* (uplifiting) and *Neubau* (new build) correlates with the building activity. This shows that the media often reports on new build projects in a neutral and positive way. Gentrification and displacement appear only rarely in connection with locations in Zurich North. The insignificant results for *Sanierung* (renovations) do not surprise since renovation projects are not reflected in the GWR dataset. Thus, they would only correlate if they were conducted at the same time as a new build project. Generally, there are two reasons why the newspaper coverage does not statistically correlate with the building activity. On one hand, the discrepancy can be caused if the media mentions a location even though there is little to no building activity. This can be explained by location name ambiguity, e.g. that the media reports on locations with the same name in other regions (example Dreispitz); reporting on projects that were completed in the past or will be constructed in the future (example Thurgauerstrasse); the referral to other locations elsewhere or reporting on these processes in general. On the other hand, the discrepancy between news coverage and building activity can be caused by the lack of media attention on building activity happening in an area, as was often the case in Affoltern and Schwamendingen.

Table 21: Results of the Spearman's' Rho analysis of the correlation between newspaper coverage and building activity between 1991 – 2024. Marked in blue are the statistically significant results (p< 0.05) whereby the shade indicates the strength (the darker, the stronger the statistical relationship. Marked in red are the results that are not statistically relevant.

Category	Frequency	Gentrifizierung	Aufwertung	Neubau	Sanierung	Verdrängung
All locations	0.11	0.08	0.09	0.19	0.07	0.07
Non-profit	-0.09	-0.05	-0.06	0.03	-0.09	-0.06
For-profit	0.16	0.12	0.13	0.23	0.11	0.11
Affoltern	0.06	0.10	0.06	0.16	0.05	0.11
Oerlikon	0.02	0.05	0.05	0.13	-0.01	0.03
Schwamendingen	0.14	0.01	0.12	0.20	0.06	0.00
Seebach	0.24	0.25	0.20	0.28	0.23	0.20

4.2 Collocation

In the following, some of the most interesting collocation results will be presented regarding the different neighbourhoods, the five keywords as well as other words of interest. The interpretation is based on the created word clouds for each of the terms and across time. An example of such a word cloud can be seen in Figure 17 while the other word clouds are located in Appendix B.

Figure 17: The top twenty collocates for the node Verdichtung (densification) for each sub-corpus (whole corpus, thematic corpus as well as different time periods) displayed as word clouds. This is an example of how the collocation results were visualised.

4.2.1 Collocations of the Neighbourhood Names

The collocation results for the different districts are quite interesting and diverse depending on the neighbourhood. While in Oerlikon and Seebach many collocates are either landmarks or place names, in Affoltern and Schwamendingen they are more about the development and history of the place. This shows that Oerlikon and Seebach are larger and perhaps more known neighbourhoods that receive news coverage about different topics while Affoltern and Schwamendingen are reported on less often and in a more specific context. Interestingly, many of the landmarks and placenames match with the building activity in these areas, supporting the results from the co-occurrence analysis. This shows that collocations in regard to district names cannot only be used to investigate the context but also find places of interests in the neighbourhoods.

Affoltern, only has results for 4 sub-corpora (whole corpus, topic corpus, 2006 – 2010 and 2011 – 2015) again highlighting that it is the neighbourhood covered least in the news. Many of the collocations

concern change (*wandel, entwicklung*), the densification (*nachverdichten*) and the population growth (*wachstum*, *rasant (*rapid*))*. The topic matches the building activity during this time, where a large areal development was constructed in the west of Affoltern that would increase the population by 25% within 5 years. Besides that, Affoltern is not really mentioned in the corpus which is highlighted also by the lack of results in other periods.

Striking about the results for Seebach are that many place and street names collocate, both within Zurich North as well as outside, which might be a bit surprising. After looking at some articles it becomes apparent that a lot of them come from articles that talk about public pools in Zurich and list different places (e.g. *Freibad* (outdoor pool), *Hallenbad* (indoor pool), *Letzigraben*, *Allenmoos*, *Altstetten* and *Mythenquai*). Other interesting collocations are *Swiss-life-Liegenschaft* and *Siedlungsrichtplan* in the result of 2021 – 2024. The reason for this is an exchange of land between the city of Zurich and the Swiss Life company, where the city of Zurich traded land in Opfikon for two strategic land reserves in Seebach (Kälin, 2022). Furthermore, many of the big areal development projects appear in the collocation results (e.g. *Köschenrüti*, *Thurgauerstrass*, *Eichrain*, *Katzenbach*, *Leutschenbach* and *Hagenholzstrasse*).

In Oerlikon, also many place names and landmarks collocate which is not surprising. Especially, the words *Bahnhof* (train station), *Messezentrum* (event hall), *Marktplatz* (marketplace), *Kantonsschule* (highschool), *Hallenbad* (indoor swimming pool), *Rennbahn*, *Neu* and *Stern* collocate often, as they are usually referred to with Oerlikon afterwards. Furthermore, places and neighbourhoods in the vicinity of Oerlikon are mentioned such as *Flughafen* (airport), *Schwamendingen*, *Seebach* and *Affoltern*. Most interesting however, are the collocates referring to the former Neu-Oerlikon region (e.g. *Maschinenfabrik*, *MFO*, *Contraves*, *Bührle*) that appear in all sub-corporas except 1991 – 1995. This again highlights the importance and wide newspaper coverage of this development area.

For Schwamendingen it is notable that there are not that many results either, having found no collocations for the period of 1991, 1996 and 2021. Comparing the results for the whole corpus and the topic corpus, they seem to be quite similar to each other, including topics such as demographics, the city planning concept of *Gartenstadt* (garden city) implemented by Steiner (*Steinerplan*) and urban development. Looking at the population many words are striking such as *Durchmischung* (mix), *dichter* (denser) and *zehnfach* (tenfold). Between 1940 and 1970 there was a tenfold increase of population due to the industrial factories in Oerlikon, as many workers moved to Schwamendingen as it was cheaper than other parts of the city. This population growth is then linked to challenges in Schwamendingen as a lot of homes had to be built in a short amount of time, which have to be renewed now as well as the high share of foreign nationals (NZZ, 2002; Feusi and Hofer, 2020). Interesting are also the many words relating to the image and character of Schwamendingen, which is said not to be the best (e.g. *Ghetto-Charakter*, 2016-2020). Furthermore, *Genossenschaft* (cooperative) is a collocate which reflects the large prevalence of cooperatives in Schwamendingen. To my surprise, big building projects such as the Mehr als Wohnen or Einhausung do not show up in the collocation analysis. The collocates and topics do not change over time but are always quite similar.

Additionally, a collocation analysis was also conducted for Leutschenbach. The results of Leutschenbach show a nice change of discourse as in the 1990s most collocates refer to the national television studio that is located there. In the early 2000, many of the collocates are on the topic of the school that will be built there and of the development area. After 2016, most collocates refer in some way or the other to developed residential estates. While Leutschenbach has similar number of articles across most time periods, Neu-Oerlikon has a peak number of articles in 2006 / 2011 where the neighbourhood was built. Afterwards, there are less articles that use Neu-Oerlikon which shows that the name did not stick to the area. A reason for this could be that the project Neu-Oerlikon is finished

for some time now while Leutschenbach is still extended. From 2001 – 2015 Neu-Oerlikon collocates with *Entwicklungsgebiet* (development area), different location such as *Max-Bill-Platz* and *MFO-Park* as well as with verbs of activities such as *entstehen* (develop) and *bauen* (build). After 2016 words such as *gelungen* (successful) and *schön* (beautiful) collocate. This implies that the finished project is seen as overall successful, and that the development is finished. The collocates for Mattenhof are quite diverse and show that the location name is not unambiguous. Besides *Schwamendingen*, *Kriens* and *Breitenrain* collocate multiple times, where other estates called Mattenhof exist. This unambiguity is a big problem for these methods as they can lead to false results since it is impossible to determine which Mattenhof they collocate to. Close reading and if possible better filtering is necessary to reduce such errors.

4.2.2 Collocations of the Five Keywords

Some of the most interesting collocates for *Gentrifizierung* (gentrifiaction) are *Fachsprache* (technical language) and *drohend* (imminent) which can be found in both the periods of 2011 and 2016. This implies that the word is not very well known to the public and thus must be explained as well as that while there is not yet evidence for gentrification, recent developments could lead towards it. This would be fitting, as in 2016 collocates relating to causes of gentrification occur such as *renovieren* (renovate) and *aufwerten* (uplifting). Regarding locations of gentrification, the following places occur: *Rosengartenstrasse*, *Langstrasse* and *Europaallee*. None of these places are located in Zurich North but in the city centre. This is not surprising, as these are places where gentrification traditionally is defined, while the outskirts of the city are places where people are pushed to. Another interesting collocate is *schleichend* (gradual) that appears in 2016, reflecting that gentrification does not occur from one day to the other but is a phenomenon that takes paces over a longer period of time.

The collocates for *Aufwertung* (uplifting) and *aufwerten* (uplift) are often connected to locations in the city of Zurich. Thereby, newspaper articles talk often about the uplifting of public spaces (*öffentlich* (public), *städtisch* (urban), *Quartier* (neighbourhoods) as well as natural spaces for biodiversity (*ökologisch* (ecological)). The number of articles containing the word have been similar over time, showing its timeless relevance. There are many collocates that can be found in different periods, indicating that the discourse around it did not fundamentally change. However, in the two most recent periods also a few negatively connotated words collocate such as *yuppisierung* and *verdrängung* (displacement) appear while in the period before words such as *Chance* (opportunity), attractive and positive appear. This could show that while overall time periods places are being uplifted, the media has become more aware of the potentially negative effect the improvement of neighbourhoods and live quality can have.

The collocations for *Neubau* (new build) are very similar across all sub-corpora which indicates that the context of the usages has not really changed over the years and is used specifically in the context of construction and building projects. Words that appear often are for example *ersetzen* (replace), *planen* (plan), *entstehen* (develop), *abreissen* (demolish), *realisieren* (implement), *abbrechen* (demolish) and *weichen* (give way) which are all very specific to construction projects as they describe that what was there before will make way to something new. Furthermore, words such as *kosten* (cost) and *teuer* (expensive) collocate often. Additionally, words referring to what is built collocate. For example: *Schulhaus* (school), *Einfamilienhaus* (single family house) or *Mehrfamilienhaus* (apartment building). Interestingly, *Schulhaus* (school) appears most often regarding the type of building that is built. One explanation is that if city projects cost more than 20 million francs there has to be a public vote on it. Since schools are large projects that cost a lot of money, newspaper inform about the project before the vote. With a few exceptions (e.g. *Leutschenbach* and *Messe*), the type of projects are general terms and not specific projects.

Sanierung (renovation) and *sanieren* (renovate) is also, unsurprisingly, used in the context of construction which is shown by the similarity of the results of the whole vs. the topic corpus. Many collocates include the completeness of the renovation (*umfassend, umfangreich, total, grundlegend*), showing that the newspaper articles mostly report on large renovations. In contrast to that, in 2006 and 2011, *sanft* (soft) also collocates with renovation, meaning that the tenants can stay in the apartment since the modifications are only minor. The usage of soft during this time could indicate that there is a discussion going on around a more socially responsible way of renovating. However, to confirm this more close reading would have to be conducted. Interesting is also that in the period of 1991 many collocates are technical and refer to some type of infrastructure compared to the collocates in other periods where they talk more about the process as well as the urgency (*dringend, notwendig*). There are a few specific projects mentioned, most interesting of all the *Schöneichtunnel*, appearing from 1996 – 2015 and the *Einhausung* which is a collocate in 2006 – 2010.

The results for *Verdrängung* (displacement) and *verdrängen* (displacing) are quite different. The collocates around the verb are very diverse, ranging from animals to football to *Gewerbe* (businesses) and demographics (*Mittelstand* (middle class), *Einkommensschicht* (income bracket), *Arbeiterbevölkerung* (working class) indicating that the word is not used in one context only but is used to describe the displacement of different objects and subjects which makes it difficult to interpret. However, some of the interesting collocates regarding urban transformation that appear from 2006 onwards, are middle class, inner city, businesses or working class, perhaps implying where it occurs and who might be affected by it. More interesting than *verdrängen* (displacing), are the results for its corresponding noun *Verdrängung* (displacement) which is clearly more often used in an urban context. Especially in recent years, from 2016 onwards, collocates such as *Gentrifizierung* (gentrification), *Luxussanierung* (luxury renovation), *Spekulation* (speculation), *Aufwertung* (uplifting), *Verdichtung* (densification), *Angst* (fear) and E*ntmischungsprozesse* (segregation processes) are found. This implies that there is an increased awareness and fear of the negative effects of uplifting, densification and speculation especially in regard to the displacement of low-income (*einkommensschwach*) households through financially stronger households. It is interesting that the collocates include causes, consequences, sentiment (fear) as well as (undefined) interventions. However, already in 1996, *Wohnraum* (living space), *zunehmend* (increasing) and *Konflikt* (conflict) appear as collocates, indicating that there was already an issue around housing spaces at that time.

4.2.3 Collocations of other Keywords of Interest

One of the more interesting collocation results are those for *Industriebrache* (industrial fallow). The first result appears in the period of 1996 where, the Maag areal located in Zurich West is mentioned. Furthermore, words such as *Umgang* (handling)*, Nutzung* (usage)*, interessiert* (interested) or *ungenutzt* (unused) imply that there are industrial sites that are unused and that there is a discussion going on about how they should be used. In the 2000s words related to *Umnutzung* (conversion) and *Entwicklung* (development) dominate together with potential kinds of conversion: residential, trade areas and shopping centre. In 2006 *Hunziker-Areal* (Leutschenbach) and *Neu-Oerlikon* collocate with *Industriebrache*, as they are prominent examples of such transformations. In 2011 *Innenenwicklung* (inward development) collocates with the industrial fallow, which makes sense as the new spatial planning act (Raumplanungsgesetz, RPG), established in 2014, enforced inward development. In 2016, *verdrängen* (displacing) collocates with *Industriebrache* (industrial fallow)which again shows that there is more awareness of the negative effects, densification and inward development can have.

For *Innenentwicklung* (inward development), there are only results in the topic, 2011 and 2016 sub-corpus, showing that it is a term that is only used recently and in a very specific context. In 2011 inward development is used with industrial fallow, densification and *Überbauung* (areal development) indicating perhaps that the newspapers explain what inward development means. In 2016, the

collocates relate to the how it should be done which is *sozialverträglich* (socially responsible) as well as *vorbildlich* (exemplary).

Most collocates of *Quartier* (neighbourhood) are positive, such as *lebendig* (lively), *lebensqualität* (life quality), *Durchmischung* (mix), booming and urban. The discourse around it has not changed much over time and the collocates are quite similar, including a lot of different neighbourhood names. After 2006, *wachsen* (growing) collocates with neighbourhood reflecting the population growth at that time. While the population decreased between 1960 and 1990, more people moved back into the cities. With the Neu-Oerlikon neighbourhood and the start of the Leutschenbach more living room was built in the northern districts thus more people moved into these areas. An interesting collocate is *beleben* (revitalising) in the sub-corpus of 2011 which could refer to Neu-Oerlikon, as the development there was criticized for being too sterile and clean with only few attractive businesses and cafés that invite people to stay.

Furthermore, the results for *Verdichtung* (densification) are interesting (more so than the results for *verdichten* (densify). Densification starts to become a relevant topic in Zurich North from 2006 since the population has grown substantially, whereby the most frequent collocate is *massvoll* (moderately), urging that it must be done carefully (Figure 17). In 2011 the collocates seem to be about successful and artful densification projects (*stark, gelungen, geschickt*). In the sub-corpus of 2016, the focus seems to shift to the importance of quality of vertical densification in the city centre. In 2021, the collocates tell a story of densification showing that it is increasingly relevant, giving reasons and effects: due to the population growth (*wachstum*) densification is needed which leads to more urbanisation and can cause displacement (*Verdrängung*). Housing cooperatives seem to be also a focus of the discourse around it, possibly referring to the large replacement construction projects in Zurich North. While with *Verdrängung* (displacement) there is a negative word in it, the densification in itself is deemed to be reasonable (*sinnvoll*) in the newspaper articles.

Last but not least, I analysed the collocates for *Wohnungsnot* (housing shortage). There are not many results, the first being in the period of 2011 where *leiden* (suffering) and *minderbemittelt* (of small means) collocate. An interpretation for these results is that the housing shortage started to become a problem but was mostly affecting low-income households. In 2021 the problem seems to have increased and is now described as *drängend* (urgent) and *anhaltend* (persistent). Furthermore, the *Leerwohnungsziffer* (share of empty apartment) collocates which is one way to measure the housing shortage. This shows that the problematic of a housing shortage is currently very relevant and that there is a high awareness of it.

4.3 Topic Modelling

Depending on the result of the best model approach, I defined the parameters and created a topic model for each of the sub-corpora. The input hyperparameters for each topic can be found in Table 22. In the following, I will analyse the topics generated in the topic models and discuss the most interesting keywords. Most strikingly, despite the different number of topics for each sub-corpora, the topics generated by the model are very similar to each other and can generally be categorized in three topics: “urban development”, “parliamentary discussions” and “police/traffic news”. This makes sense in so far that the keywords are all related to each other in some ways and that the media has some general topics they report on. Additionally, some of the topics cannot be labelled since the keywords are too diverse and for the sub-corpus *Verdrängung* (displacement), the category “football” can be made out as well. Furthermore, for the gentrification sub-corpus a Large Language Model was also used besides the LDA, to find the top twenty topics in the corpus and then compare the results. The LDA and LLM outputs were quite similar as both contained terms relating to urban development as well as politics.

However, the LDA results contained more specific words such as cooperatives and Europaallee and thus allowed more detailed insight into the corpus than the LLM results which were quite broad and could be divided in "Urban development", "general newspaper categories" and "various" (Table 23). The similarity of the results can partially be explained by the low number of articles in some sub-corpora (e.g. *Gentrifizierung, Verdrängung*), the similar context of the keywords as well as the occurrence of several keywords in the same articles. Because the different topic model results are quite similar, I will discuss each of the three topics individually based on the example of *Aufwertung* (uplifting) instead of the topic model output for each of the sub-corpora, which can be found in Appendix C.

Table 22: Hyperparameter for each topic model as determined by the grid search that chose the model with the highest log likelihood and the lowest perplexity.

Sub-corpus	Prior of document topic distribution	Learning decay	Learning offset	Number of topics
Gentrifizierung (gentrification)	0.1	0.6	1000	5
Aufwertung (uplifting)	0.1	0.7	10	5
Neubau (new build)	0.1	0.7	100	7
Sanierung (renovation)	0.01	0.6	1000	9
Verdrängung (displacement)	1	0.7	10	5

Table 23: Results of the LLM for the sub-corpus of articles containing gentrification divided by the three detected categories "Urban Development", "Newspaper Categories" and "Various"

Urban Development	Newspaper categories	Various
Alt (old)	*Gesellschaft* (society)	*Familie* (family)
Architektur (architecture)	*Kultur* (culture)	*Fernsehen* (television)
Bau (construction)	*Natur* (nature)	*Freiheit* (freedom)
Wohnen (living)	*Politik* (politics)	*Politiker* (politicians)
	Wirtschaft (economy)	

The five different topics that were modelled for the sub-corpus *Aufwertung* (uplifting) can be categorised in three distinct categories and one that cannot be labelled: police news (Topic 1), urban development (Topic 2 and 3) and political process (Topic 4) (Figure 18). Topic 1 contains words such as *Kantonspolizei* (cantonal police), *Sachschaden* (property damage), *Spital* (hospital), *Fahrzeug* (vehicle) or *Medienmitteilung* (press release) which leads to the conclusion that a share of newspaper articles in the corpus contains press releases from the police that inform about accidents and other incidents. The most likely explanation for this result in the uplifting sub-corpus is the prevalence of tickers and short notices that contain a variety of different information ranging from police news to small infrastructure projects to information about vote outcomes. Topics 2 and 3 contain many words such as neighbourhood, train station or space, reflecting different things that can be uplifted in an urban context. Furthermore, the terms *Neubau* (new build) and *Genossenschaft* (cooperative) can be found in the results which shows how relevant they both are in uplifting Zurich's neighbourhoods. Topic number 4 contains many terms relating to the political processes and levels of government such as *Parliament, Stadt-, Gemeinde-* and *Regierungsrat* (parliament, city-, municipal, state council) or *Postulat, Vorlage, Antrag* (postulate, bill, motion) which shows that the government and especially local councils take a large part in the city development. In the other topic modelling outputs, also terms such as costs, schools and initiative are found in this category as many uplifting projects especially for public spaces and infrastructure (such as tunnels, schools etc.) are done by the government and thus are discussed in parliament. Furthermore, if the cost of a project surpasses a certain amount, there has to be a public vote on it. For public votes there is always a lot of newspaper coverage both before the

vote to inform the public about the project / initiative as well as afterwards to summarise and discuss the results. Informing the public about discussions in the parliaments and votes is an important part of daily newspaper work and therefore it makes sense that it is reflected in the topic model. Lastly, topic 5 contains a variety of different, unrelated terms which makes it very difficult to label. This topic demonstrates nicely the limitations of topic modelling as it is only useful as long as the reader can make sense of it and enrich it with background knowledge.

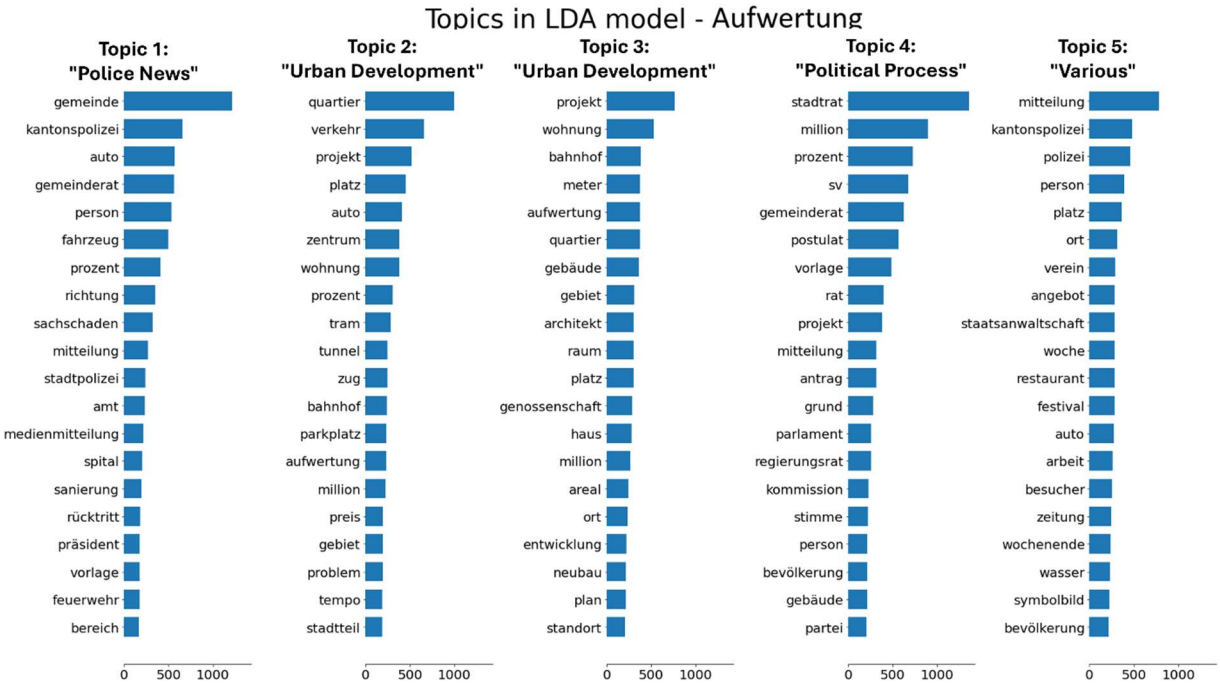

Figure 18: Results of the topic model for the uplifting sub-corpus which will stand in as an example for the other sub-corpus as well. The five topics can be categories into three distinct categories: police news, urban development and political process. Due to its large variety of keywords, Topic 5 cannot be categorised.

5. Discussion

The objective of this chapter is to discuss and answer the two research questions presented in section 1.1 in the context of the literature and the results presented in the previous chapters. Furthermore, the methodology used in this thesis as well as the limitation will be critically discussed. Last but not least, an outlook to further research will be given.

5.1 Using NLP Methods to Study Urban Spaces

Research question 1: In what ways can NLP tools be used to analyse discourse around gentrification and displacement over time in a newspaper corpus?

Natural language processing methods are useful to analyse transformation and gentrification processes in urban spaces based on a newspaper corpus and allow to identify interesting articles, spaces and discourses. This in line with the findings of Jiang et al. (2023) that applied Natural Language Processing (NLP) tools to a newspaper corpus as well to analyse the phenomenon of urban shrinkage. However, NLP methods are very diverse and can be conducted in many ways. This is why it is important to focus on a few methods and carry them out diligently and in a thoughtful manner while having the main goal and research question always in mind. Otherwise, there is a big risk of getting lost, experimenting with different approaches and adjusting models. In this thesis, the focus was on three different tools that were used to analyse a newspaper corpus about the Zurich North region that spans over the last thirty years in regard to how they report on gentrification and urban development. Each method looks at a different aspects of the corpus and allows to recognise certain patterns.

The frequency and co-occurrence analysis for example enabled to see what locations featured prominently in the corpus at different points in time and in which context. Interestingly, a handful locations such as the Leutschenbach or Neu-Oerlikon had high frequencies over almost the entire period from 1991 to 2024. This makes sense as they are names for large projects that transformed former industrial areas into mixed-use residential neighbourhoods. Other locations on the other hand only had high frequency in certain periods, such as the Blumenfeld in 2011, which were usually related to an incident or a certain project, while almost half of the locations in the gazetteer never occurred in the corpus. To differentiate between places that occur frequently in the corpus in general and those that occur frequently regarding urban change, the comparison between the frequency and co-occurrence results was very useful. While the frequency and co-occurrence analysis on their own already gave some insight into interesting places, it was especially the comparison with the building activity that allowed to draw some more substantial conclusions. The construction activity and news coverage correlated partly in a statistically significant way, most often for new build and the Leutschenbach. Thereby, both, places where the coverage matched the construction activity as well as where it did not match were interesting as they served as an indication for close. Especially the created maps were useful as they allowed to easily spot locations of interest and then filter the according articles by looking for the keyword and location name. For the map, the building activity was displayed weighted by building volume to highlight large buildings and towers over smaller constructions. Generally, this worked well but, in some cases, where single large buildings were constructed such as warehouses or event halls, the weighting by volume could lead to false impressions as these constructions do not really uplift a neighbourhood or transform it, especially if they are built in an industrial area. To avoid such confusion, industrial buildings could have been excluded from the analysis, a differentiation between residential and commercial buildings could have been done or the weighting chosen differently.

While the frequency and co-occurrence analysis allowed to find locations of interest and understand in what context they are used in, the collocation analysis allowed to understand how and in what context different terms are being used in the news and how this changed over time. Based on the collocates of the node, the context of the usage could be interpreted. However, depending on the node this was either very intuitive or almost impossible. This depended mostly on the variety of articles and context the word was used in. Another reason for the diversity of collocates with certain keywords was the inclusion of news ticker and other short news formats in the analysis. Because in these formats they merge different information and news that are not related at all. Collocation, however, makes use of the structure of text and that the proximity of words with each other means that there is a relation. With these tickers this is not the case and therefore this could falsify the results. Generally, the results for the collocation analysis are sometimes a bit of a black box as quite a few of the results contain spelling errors, are very specific or oddly composed nouns that are not normal in everyday language. It is thus hard to imagine that these words would collocate often with a node across the whole corpus. Possibly, these results are artefacts of the duplication of some articles. Another possibility could be that the node does not appear often in general, resulting in rather odd words collocating. To reduce this problem, a larger collocation window could be considered as this increases the chance that certain words are found within the proximity of the node multiple times. Furthermore, the results could be assessed statistically. By conducting the collocation analysis for the different time periods, I was also able to detect some changes in the discourse. However, for many of the nodes the discourse did not change significantly over time. Nonetheless, the collocation analysis helped to identify words of interest for different periods which then could be analysed in more detail by close reading. Furthermore, the results indicated which words are used only in the context of urban development and which words are used more broadly and thus might be less relevant to analyse.

Finally, I will take a look at the topic modelling outputs. Perhaps the biggest added value, a topic model can bring is that it shows in what different context a word is used in a big corpus. This can help to determine important keywords in the beginning and help to avoid misunderstandings. For example, the usage of *Verdrängung* (displacement) in the context of football or the large presence of police statements and traffic accidents in general could have helped to early identify that these types of articles are irrelevant for the analysis. Thus, these articles could have been excluded from the corpus before the co-occurrence and collocation analysis, which would have improved the results. While this would be a very helpful approach, especially in the beginning of an analysis, topic modelling itself does not really reveal new insights for the context of certain keywords compared to the collocation results, as the outputs are very broad and generic. This makes it difficult to understand the context in more detail and a lot of close reading is still required when working with topic modelling. However, it does give a hint in what ways a word is used, allowing to analyse how well a keyword is suited for the topic analysis. Another caveat of the topic modelling is the long time it takes to compute and generate results. This makes it difficult to use in an explorative way especially for big corpora. To improve the quality of the topic modelling as well as efficiency, some measures could be taken. For example, an exclusivity and coherence plot could help to identify the ideal number of topics for each of the sub-corpora, to avoid a large overlap of the words and make them more legible. Furthermore, embedding models could be used to capture the semantic relationship between words and sentences which has been shown to improve the topic modelling results (Jiang *et al.*, 2023).

In summary, the three NLP methods used in this thesis are all relatively simple to implement and allow to navigate through a large corpus of newspaper articles that a person on its own could not have handled within a reasonable amount of time. The three methods enable to identify interesting locations and terms related to the topic of interest. In this thesis, the three methods were applied simultaneously to explore what each of the different methods can reveal. In hindsight however, a

subsequent approach probably would have been better. In a first step, a topic modelling could have been conducted to explore the corpus in broad way and the different keyword sub-corpora. Based on the topic model output, the corpus could have then been filtered more thoroughly in the pre-processing to increase meaningfulness and relevance of the results. For example, the high prevalence of articles about football or police notifications could have been detected early on and consequently be excluded from the corpus before conducting a co-occurrence and collocation analysis. Furthermore, all three methods used, are based on a bag-of-word model and thus are highly simplistic and neglect the context of the articles. A good knowledge about the subject matter and the research area are therefore essential to interpret the results. Thereby, the results should always be complemented with close reading to gain a deeper understanding of the subject matter and assess the results. These conclusions are in line with what Purves, Koblet and Adams (2022a) have found.

5.2 Newspaper Coverage on Gentrification and Urban Development in Zurich North

Research question 2: How do newspaper talk about gentrification and urban development in the northern districts of Zurich?

5.2.1 Newspaper coverage on gentrification and urban development

Newspapers report to some extent on urban development in the region Zurich North. Thereby, they often do so in a neutral and informative fashion which manifests in the high occurrence of *Neubau* (new build) and *Sanierung* (renovation) that co-occur the most with different locations. This makes sense in so far as they inform about certain projects or also possible bypasses that occur due to construction. Even though *Neubau* and *Sanierung* co-occur by far the most, newspaper articles also often use *Aufwertung* (uplifiting) to refer to building projects which shows that they see new constructions in a mostly positive light. The more negative aspects such as *Gentrifizierung* (gentrification) and *Verdrängung* (displacement) on the other hand occurred only rarely in newspaper articles of the past thirty years, which shows that the negative effects are not often talked about regarding Zurich North. Thus, the media contributes only little to raise awareness on the issue. Gentrification first appears in 2006 and is almost exclusively used in relation to the Langstrasse and Europaallee in the centre of Zurich and mostly brought up by experts (sociologists and urban geographers) or activists (from political parties, neighbourhood lobbies etc.). Interestingly, the Seefeld does not occur in any of the results even though it has become a symbol of gentrification in Zurich.

There are three possible explanations on why gentrification does not occur in the corpus that often. First of all, gentrification is an originally English and technical term with a fuzzy concept that might not be very well known in in the German speaking public. While experts and activists are familiar with it and use it in few instances, the broader public might be less familiar with it. Since newspapers write for the broad public, they might refrain from using *Gentrifizieren* and instead write differently about it. However, since *Verdrängung* (displacement), which is closely related to the issue of gentrification, is also not used very often in the corpus and in relation to people, more research would have to be conducted to find out if and how newspapers write about gentrification instead. The second explanation has to do with the location of Zurich North and the starting point of the transformation. Zurich North is located on the outskirts of Zurich and has only started transforming in the last twenty years with many of the larger transformations taking place on industrial brown sites. Extensive replacement constructions, that lead to the direct displacement of people, have only recently set in. Because gentrification is traditionally understood as a process taking place in the inner city that directly displaces people (Slater, 2011), journalists might not have it on their radar that the process can also occur in other places than the inner city and through the general uplifting of an area. Additionally, the transformation of Zurich North is still ongoing, and many effects are only unfolding just now. To

investigate whether gentrification is not used in Swiss newspapers in general or whether it is simply not used in relation to Zurich North, a general corpus about gentrification in Switzerland should be gathered and analysed. Evidence that the news does use the term but not in relation to Zurich North is that a query for articles containing Langstrasse, Zürich and gentrification with the Swissdox@LiRi resulted in 274 articles from the period 1991 – March 2024. This indicates that the awareness around the issue of gentrification exists mostly in a few stereotypical inner-city hotspots. To confirm this however, more research should be conducted around the newspaper coverage in other regions. A third explanation for the few occurrences of gentrification in Zurich North is that because there are many housing cooperatives active in Zurich North, the issues of uplifting, renewal and renovation are perceived as less pressing since they ensure lower rents and a degree of social mixing. This reasoning was especially expressed in recent articles relating to the effects of the highway enclosure in Schwamendingen. However, it must be kept in mind that even though the rents of housing cooperatives are lower than those on the private market, they are still increased when replacing old buildings which can displace former inhabitants. Since the population is expected to grow further over the next years especially in Zurich North, more attention to these negative issues should thus be paid in the future (City of Zurich, 2024b).

5.2.2 Reporting on locations compared to the building activity

The newspaper coverage was in most periods not statistically significantly correlated with the building activity at that time. There are different explanations for why the peak of news coverage does not match the construction phase which have mostly to do with certain aspects of newspapers and media that must be considered when using them as data source. First of all, newspapers report on issues that are newsworthy and currently relevant. Building projects, and especially large ones can take a long time to implement. Thereby, the discussions around the project and potential issues can precede the five-year time periods that were analysed in this thesis. Thus, newspapers might report on a project only as long as the discussions are going on. As soon as the construction starts, the issues are settled and there is nothing worthwhile to report on anymore, unless there are complications. This thesis, possibly, did not consider this gap between discussion and constructions enough.

Secondly, newspapers often report on issues that their readers are interested in. This leads to topics having cycles of more representation when the topic is new, and readers are interested in and less representation when the readers have become oversaturated. This could explain the lack of coverage on the areal development in Affoltern in 2006 as it occurred between the transformation of Neu-Oerlikon and Leutschenbach that caught a lot of media attention. Thirdly, newspapers can include a certain bias due to who writes the articles and who they write it for. According to a study, 70% of the journalists in Switzerland have a higher education degree (Dietrich-Gsenger and Seethaler, 2019) and thus are likely to be part of higher and middle socio-economic groups. Assumingly, many of the journalists as well as their target audience live and move around predominantly in the city centres, they have less interest and awareness on topics and processes occurring on the city fringes. Especially since gentrification and urban transformations are processes that unfold over a long period of time, one would need to observe it for a longer time to become aware of it or have a loud group of activists who draw attention to the problem. This could be a reason as well why especially Affoltern received only little attention over the analysed period.

Fourthly, the newspapers in the corpus were predominantly national papers while the topic analysed in this thesis is local. In the national papers there are usually only few pages dedicated to the Zurich region, thus journalists would have to set priorities what they report on. For future analyses it is therefore recommendable to match the newspaper sources to the topic of analysis. For example, local newspapers should be considered when studying local areas. However, as seen with the Zurich Nord newspapers, it can be difficult to gather enough data, as small local newspapers might lack a digital

archive which can be accessed easily. Last but not least, gentrification takes place in a bigger societal context between the necessity of densification and the discourse around who can live in a city (see the debate on the right to the city (Schmid, 2012; Purcell, 2013)). Therefore, newspaper might report on the issue not in relation to individual building projects at the time it is built or certain streets but as a general trend and issue.

Even though, the correlation between the newspaper coverage and building activity are mostly not significant, there are a few points in time where they correlate in a significant, positive way. Especially, the *Neubau* (new build) co-occurrence results as well as the locations in Seebach correlated positively in different time periods as well as over the whole period. The collocation between *Neubau* and the building activity make a lot of since it is directly connected with each other. Thereby, the newspapers inform on different buildings that were completed or will be completed soon. Compared to *Neubau*, *Sanierung* (renovation) only correlates significantly in a few instances. This makes sense as renovations are not included in the Swiss Federal Housing and Building Register (GWR) dataset and thus reports on renovations that do not happen at the same time as a new builds such as tunnels or infrastructure will not match with the GWR data. The correlation between the building activity and Seebach are most likely caused by the development in the Leutschenbach area. The development in this area has already started around 2000 and often included large building complex or towers with a high volume. Because the building activity was weighted by volume, reports in this area would have a bigger impact on the correlation analysis and comparison than for example smaller buildings.

When comparing the newspaper coverage with the building activity it also stands out that a few places are reported on very frequently in all periods despite the actual building activity. Examples are the Leutschenbach, Neu-Oerlikon or Einhausung Schwamendingen and the streets Schaffhauserstrasse, Wehntalerstrasse and Winterthurerstrasse which appear in the top five of many categories and periods. The high frequencies of Schaffhauserstrasse, Wehntalerstrasse and Winterthurerstrasse are most likely related due to their length and relevance in the region even though there were no single large construction projects conducted. Leutschenbach, Neu-Oerlikon and Einhausung on the other hand are all large construction projects. Due to their size, the planning of such projects takes more than five years and often includes a design plan or masterplan. These are usually approved by regional councils and a referendum can be held against them or objections made. If a referendum is held, there is a vote (as for example was the case of Thurgauerstrasse,), which is then reported in the newspapers. These processes can take many years, which is why they appear in the newspaper corpus long before the actual buildings are constructed. Generally, it is striking that many locations where there was a vote on have high frequencies in the corpus. The high frequencies can be explained by the exhaustive media coverage on the different initiatives and proposals in the run-up as well as the reports on the results. This is especially true for different school projects such as Birch, Blumenfeld or Leutschenbach. Due to the large population growth in Zurich North, many new schools were built during the past decade and others are still in the planning. Schools are expensive public infrastructure projects that usually require a public vote to approve of the costs. Often, there are also controversial disputes around these high costs which is why there is a lot of newspaper coverage around it (e.g. the dispute around the Blumenfeld in Affoltern, which was the most expensive school in the city at that time (Huber, 2012; Troxler, 2013) or the current discussion around the new school in Saatlen, Schwamendingen (SRF, 2021; von Ledebur, 2023)). This is of interest, as schools are an indicator for population growth in an area which in the case of Zurich North is caused by densification measures and areal developments. While newspapers do not seem to report on individual building and densification projects, the large coverage on school projects reflects this change.

5.2.3 Change of newspaper reporting over time

Over the past thirty years, newspapers have repeatedly reported on urban development in Zurich North. The reporting on new builds, renovations and uplifting were prominent and constant across all time periods and did not change much over time. This was shown in the similarity of the collocation results across different sub-corpora. This makes sense as newspapers inform the public about individual construction, renovation and uplifting processes mostly in a neutral way such as project type, size or costs. Furthermore, the coverage on Oerlikon, Seebach and neighbourhoods in general remained mostly the same across all time periods, reporting on a variety of topics and the general liveability of the neighbourhoods. Compared to that, the reporting on Affoltern and Schwamendingen was only sporadic and limited to urban development and population growth. The collocation analysis further showed that while a lot of the chosen keywords were reported on constantly and in a similar way over time, others only emerged in recent years. For example, *Innenentwicklung* (inward development) emerged only after 2011 as it became more prominent due to the new spatial planning act, established in 2014 (Raumplangungsgesetz, RPG); *Verdichtung* (densification) was only used after 2006 as it started to become a necessity due to the urban population growth; *Gentrifizierung* (gentrification) only appeared more frequently after 2011, and *Wohnungsnot* (housing shortage) only appeared in 2011 and 2021 (this supports the observation of Stahel (2006) that housing shortage in Zurich becomes an existential problem in a regular 10-years cycle). In summary, while most of the reporting on urban change remained relatively constant over time focusing on the neutral to positive aspects of it, awareness of the negative effects of densification, new builds and uplifting seem to have grown in the past ten years. This is reflected in the emergence of negatively connotated words in the collocation analysis and the increased use of *Verdrängung* (displacement) in relation to low-income households.

5.3 Limitations

One of the biggest limitations in this thesis is the presence of duplicates in the dataset that could not be eliminated by removing all duplicate content-IDs. Since the methods applied in this thesis all use word frequencies to some extent, the duplication of content leads to a bias towards locations and words mentioned in these articles even though they might be less important. The duplication of articles occurred because of the publication of the same article in print and online media, the prevalence of news tickers and short notices that contain the same story multiple times as well as the sharing of the same article content across different media belonging to the same parent company. Reducing the duplication of content is very important in computational analyses as it can have a heavy impact on the results. However, removing duplicates is not trivial as sometimes neither the headline, publication date nor the content are identical. To minimise data duplication, different measures could be taken. On one hand, text formats such as tickers or short notices could be removed from the corpus. On the other hand, for data periods where online newspapers are already a big thing, a decision could be made to only utilise either data from print or only data from online sources. This assumes that articles are usually published in both and thus that the inclusion of both would lead to a duplication. Furthermore, texts could be compared with each other and if they have a certain degree of similarity, all but one should be excluded. Avoiding data duplication is key to improve the quality of the results. This leads to a general problem of NLP methods as the results are heavily influenced by the chosen input parameters and pre-processing steps (Purves, Koblet and Adams, 2022a). Evaluating different input parameters and pre-processing steps such as for example the statistic measure in the collocation analysis, the inputs for the topic model or another lemmatisation package could thus increase the robustness of the results.

The corpus was built based on a location gazetteer in Zurich North that was compiled with different data sources. However, it cannot be said with certainty that the gazetteer was exhaustive. For example, this thesis did not consider vernacular place names although they are important in the daily life of

many people (Davies *et al.*, 2009). Therefore, it could be that some articles about Zurich North were missed that included only these vernacular place names. The same issue is also true for the co-occurrence analysis that was based on the same gazetteer. Furthermore, a few places in the gazetteer, turned out to be ambiguous which can be a challenge for computational text analysis (Purves *et al.*, 2018) . Because bag-of-word models do not consider context, they cannot distinguish between places located in the research area and other locations with the same name outside of it. While I tried to reduce location ambiguity by including the must criteria “Zurich”, “Affoltern”, “Oerlikon”, “Schwamendingen” or “Seebach” to only find articles that related to the research area, it became apparent that it did not work entirely. Especially, the locations Dreispitz, Mattenhof, and Messegelände turned out to be ambiguous, as locations outside of Zurich North with that name featured prominently in some articles. These places could be spotted however, based on the comparison with the construction activity and some close reading, which again shows how essential ground truth data is when working with NLP tools. To minimise the ambiguity of locations, articles from newspapers outside of Zurich (such as the Basler Zeitung, Luzerner Zeitung or St. Galler Tagblatt) could have been excluded from the corpus as it is unlikely that papers from these regions report on urban transformation projects in Zurich North. Furthermore, the must criteria could have been limited to just the neighbourhood names and not “Zurich” in general as “Zurich” is used in many articles about other regions in Switzerland and beyond.

Besides the limitations of the data sources and gazetteer used in this thesis, there are also some limitations that come from the application of NLP tools in general as well as from the way they were applied in this thesis. First of all, bag-of-word models ignore the sentence structure which can lead to a misinterpretation of text due to negations, omissions or wrong referrals. This was for example the case with many of the gentrification results that either did not refer to locations in Zurich North or were negated in the articles. These limitations were also identified in other studies such as Jones, Maynard and Wartmann (2022) or Luria and Campos (2022). Secondly, the implementation of the co-occurrence analysis in this thesis partly also led to misinterpretations. Co-occurrence was defined as a keyword and location occurring in the same article, independent from the distance between these two words. This definition most likely was too wide especially for long articles. To reduce false positive results for the co-occurrence analysis in future work, the distance between two words of interest could be considered by setting a limit of how many words should maximally be between them. Last but not least, this thesis showed that the results of the different NLP methods can only be interpreted in a meaningful way if they are complemented with close reading of different texts and an understanding of the subject and research area. Otherwise, the interpretation is prone to speculations and errors due to the reasons mentioned above. This finding aligns with numerous other studies that have employed NLP methods to address a broad spectrum of questions that emphasise the critical importance of a multiscalar approach that simultaneously uses both micro and macro reading (Cai, 2021; Jones, Maynard and Wartmann, 2022; Luria and Campos, 2022; Purves, Koblet and Adams, 2022c; Taylor and Adams, 2022).

5.4 Implications for Urban Studies

This master's thesis has demonstrated the utility of NLP tools in analysing urban spaces over time, showing that such tools do not need to be overly complex. These findings are consistent with previous research (e.g. Luria and Campos, 2022; Jiang et al., 2023). While many studies leveraging NLP methods to analyse urban spaces have relied on social media data (Cai, 2021), this thesis instead utilized newspaper articles as a data source. The analysis revealed that newspapers offer a valuable alternative, as they frequently report on urban processes and societal issues while enabling longitudinal studies that extend beyond the temporal limits of social media data. With the ongoing digitization of newspapers, it is conceivable that future studies could analyse historical data over even longer timeframes. Additionally, using newspapers as a data source allows for scalability in both time and space, provided access to a digitized news archive. The selection of newspaper type should thereby align with the spatial resolution of the research question. To validate the results of NLP analysis and identify both areas of interest and media blind spots, these findings can be supplemented with data on actual building activity and close reading insights. In conclusion, NLP tools in combination with digitized newspaper archives offer promising opportunities for urban researchers, as they are powerful to address a wide range of research questions related to urban spaces. Because of their simplicity and language independence the methodology applied in this study could be applied to other neighbourhoods, cities, or countries and topics such as housing shortage and densification, enabling comparative analyses and more insights into how urban processes are perceived over time.

6. Conclusion and Future Work

In this study, almost 45’000 newspaper articles were studied by computational text analysis to find out how suitable Natural Language Processing (NLP) tools are to analyse urban issues as well as how the media reported on urban transformation and gentrification processes in Zurich North between 1991 and 2024.

This thesis demonstrated that (1) newspaper articles are interesting sources to analyse urban issues; (2) newspapers did report on urban development in Zurich North relatively steadily over the past 30 years whereby large development areas such as Neu-Oerlikon and Leutschenbach as well as projects that required a public vote (e.g. schools and layout plans) were covered the most; (3) the newspaper coverage partially correlates in a significant positive way with the actual building activity; (4) while there was a lot of coverage on renovation and new builds and some on uplifting over the years, there was only little to no coverage on gentrification and displacement.

Summing up, this thesis showed that NLP-Tools tools are suitable to analyse large quantities of text and that relatively simple methods such as frequency, co-occurrence and collocation are useful to gain exciting insights into discourse over time, while also highlighting the importance of expert knowledge, close reading and data pre-processing.

It has also become evident, that the media do only little to raise awareness of the negative effects densification projects can have in Zurich North. However, paying close attention to these issues is critical to take early measures and protect vulnerable persons in our society that are at most risk of being affected by displacement.

Based on this thesis there are different paths imaginable for further studies. One focus for example could be to extend the NLP analysis on other genres of text such as city council protocols, tweets or annual reports of housing cooperatives and property developers. Thereby, it could be analysed if they have a higher awareness of gentrification and displacement processes in Zurich North than newspapers and how it changed over time. Additionally, it would be interesting to explore if local, regional and national newspapers report differently on urban transformation and gentrification or analyse the impact the political orientation of a newspaper has on the coverage on urban issues. Another interesting direction for future work would be to analyse the newspaper coverage on gentrification in other cities and countries. It would be particularly interesting to compare news coverage within a city to identify differences regarding for example location (inner-city vs. outskirt) and social status of a neighbourhood. Furthermore, a cross-country comparative study would be exciting to analyse how the newspaper coverage on gentrification differs depending on geographical region, city size or language spoken.

7. Literature

Aalbers, M.B. (2019) 'Introduction To The Forum: From Third To Fifth-Wave Gentrification', *Tijdschrift voor Economische en Sociale Geografie*, 110(1), pp. 1–11. Available at: https://doi.org/10.1111/tesg.12332.

ABZ (2024) *Ruggächern*. Available at: https://www.abz.ch/bauten/siedlungen/ruggaechern/ (Accessed: 12 July 2024).

Architonic (2024) *Hochhausensemble Hagenholzstraße*. Available at: https://www.architonic.com/de/project/max-dudler-hochhausensemble-hagenholzstrasse/5104250 (Accessed: 11 July 2024).

Atkinson, R. (2015) 'Losing One's Place: Narratives of Neighbourhood Change, Market Injustice and Symbolic Displacement', *Housing, Theory and Society*, 32(4), pp. 373–388. Available at: https://doi.org/10.1080/14036096.2015.1053980.

Brezina, V. (2018) 'Semantics and Discourse', in *Statistics in Corpus Linguistics A Practical Guide*. Cambridge University Press, pp. 66–101. Available at: https://doi.org/10.1017/9781316410899.004.

Brown-Saracino, J. and Rumpf, C. (2011) 'Diverse Imageries of Gentrification: Evidence from Newspaper Coverage in Seven U.S. Cities, 1986–2006', *Journal of Urban Affairs*, 33(3), pp. 289–315. Available at: https://doi.org/10.1111/J.1467-9906.2011.00552.X.

Brusa, N. (2017) 'Die längste Strasse Zürichs in 45 Sekunden', *Tagesanzeiger*, 19 October. Available at: https://www.tagesanzeiger.ch/die-laengste-strasse-zuerichs-in-45-sekunden-608958114297 (Accessed: 4 July 2024).

Bürgi, L. (2014) *Gentrification im Zürcher Langstrassenquartier zwischen 1994-2014*. Master Thesis. Universität Zürich.

Cai, M. (2021) 'Natural language processing for urban research: A systematic review', *Heliyon*. Elsevier Ltd. Available at: https://doi.org/10.1016/j.heliyon.2021.e06322.

Canton of Zurich (2024) *Wohnbauten gemeinnütziger Bauträger*, *Geoportal*. Available at: https://geo.zh.ch/data/maps/59062f32-0542-4d82-ab7a-49e910a6716e (Accessed: 4 September 2024).

Cavicchia, R. (2023) 'Urban densification and exclusionary pressure: emerging patterns of gentrification in Oslo', *Urban Geography*, 44(7), pp. 1474–1496. Available at: https://doi.org/10.1080/02723638.2022.2100174.

Chen, Y., Silva, E.A. and Reis, J.P. (2021) 'Measuring policy debate in a regrowing city by sentiment analysis using online media data: A case study of Leipzig 2030', *Regional Science Policy and Practice*, 13(3), pp. 675–692. Available at: https://doi.org/10.1111/rsp3.12292.

City of Zurich (2000) *Entwicklungsskonzept Leutschenbach*. Zürich. Available at: https://www.stadt-zuerich.ch/hbd/de/index/staedtebau/planung/entwicklungsgebiete/leutschenbach/veranstaltungen/entwicklungskonzept_2000.html (Accessed: 5 July 2024).

City of Zurich (2014) *Home alone - Einpersonenhaushalte in der Stadt Zürich*. Zurich. Available at: https://www.stadt-zuerich.ch/prd/de/index/statistik/publikationen-angebote/publikationen/webartikel/2014-09-29_Home-alone_Einpersonenhaushalte-in-der-Stadt-Zuerich.html (Accessed: 1 March 2024).

City of Zurich (2023a) *Quartierspiegel Affoltern 2023*. Zurich.

City of Zurich (2023b) *Quartierspiegel Oerlikon 2023*. Zürich.

City of Zurich (2023c) *Quartierspiegel Saatlen 2023*. Zürich.

City of Zurich (2023d) *Quartierspiegel Seebach 2023*. Zürich.

City of Zurich (2023e) *Wohnfläche pro Person (m2) und Personen pro Wohnung*. Zurich.

City of Zurich (2024a) *Bevölkerungsbestand und -entwicklung*. Available at: https://www.stadt-zuerich.ch/prd/de/index/statistik/themen/bevoelkerung/bevoelkerungsentwicklung/aktueller-bevoelkerungsbestand.html (Accessed: 29 February 2024).

City of Zurich (2024b) *Bevölkerungsszenarien*, *Präsidialdepartement*. Available at: https://www.stadt-zuerich.ch/prd/de/index/statistik/themen/bevoelkerung/bevoelkerungsentwicklung/bevoelkerungsszenarien.html (Accessed: 17 September 2024).

City of Zurich (2024c) *Entwicklungskonzept/Leitbild*, *Hochbaudepartement*. Available at: https://www.stadt-zuerich.ch/hbd/de/index/staedtebau/planung/entwicklungsgebiete/leutschenbach/leitbild_konzept.html (Accessed: 4 September 2024).

City of Zurich (2024d) *Neubau Gesundheitszentrum für das Alter und Wohnsiedlung Eichrain*. Zürich. Available at: https://www.stadt-zuerich.ch/hbd/de/index/hochbau/bauten/bauten-realisiert/alterszentrum-wohnsiedlung-eichrain.html (Accessed: 2 August 2024).

City of Zurich (2024e) *Neuer Rahmen um die Genossenschaftssiedlung am Katzenbach*. Available at: https://www.stadt-zuerich.ch/hbd/de/index/ueber_das_departement/medien/medienmitteilungen/2010/maerz/100312a0.html (Accessed: 12 July 2024).

City of Zurich (2024f) *Neu-Oerlikon*. Zürich. Available at: https://www.stadt-zuerich.ch/hbd/de/index/staedtebau/planung/entwicklungsgebiete/neu-oerlikon.html (Accessed: 5 July 2024).

City of Zurich (2024g) *Quadro*. Zürich. Available at: https://www.stadt-zuerich.ch/hbd/de/index/staedtebau/planung/entwicklungsgebiete/leutschenbach/projekte_realisiert/quadro.html (Accessed: 11 July 2024).

City of Zurich (2024h) *Quartierspiegel - Über 800 Seiten Information zu Zürichs Quartieren*, *Quartierspiegel*. Available at: https://www.stadt-zuerich.ch/prd/de/index/statistik/publikationen-angebote/publikationen/Quartierspiegel.html (Accessed: 22 February 2024).

City of Zurich (2024i) *Strassennamenverzeichnis*. Available at: https://www.stadt-zuerich.ch/geodaten/download/Strassennamenverzeichnis (Accessed: 2 September 2024).

City of Zurich (2024j) *Weiterhin hohe Wohnbautätigkeit*. Zürich. Available at: https://www.stadt-zuerich.ch/prd/de/index/statistik/publikationen-angebote/publikationen/webartikel/2024-02-06_Weiterhin-hohe-Wohnbautaetigkeit.html (Accessed: 28 August 2024).

City of Zurich (2024k) *Weiterhin kaum leere Wohnungen in der Stadt Zürich*, *Stadt Zürich*. Zürich. Available at: https://www.stadt-zuerich.ch/prd/de/index/ueber_das_departement/medien/medienmitteilungen/2024/08/weiterhin-kaum-leere-wohnungen-in-der-stadt-zuerich.html (Accessed: 28 August 2024).

Cohen, J. (1992) 'Statistical Power Analysis', *Current Directions in Psychological Science*, 1(3), pp. 98–101. Available at: https://doi.org/10.1111/1467-8721.EP10768783/ASSET/1467-8721.EP10768783.FP.PNG_V03.

Craviolini, C., Heye, C. and Odermatt, A. (2008) *Das Langstrassenquartier: Veränderungen, Einflüsse, Einschätzungen – 1990 bis 2007*. Zurich. Available at: https://doi.org/https://doi.org/10.5167/uzh-6382.

Davidson, B.I., Hinds, J. and Racek, D. (2023) *Shifting landscapes of social media data for research*, *Times Higher Education*. Available at: https://www.timeshighereducation.com/campus/shifting-landscapes-social-media-data-research (Accessed: 30 August 2024).

Davidson, M. and Lees, L. (2005) 'New-Build "Gentrification" and London's Riverside Renaissance', *Environment and Planning A: Economy and Space*, 37(7), pp. 1165–1190. Available at: https://doi.org/10.1068/A3739.

Davidson, M. and Lees, L. (2010) 'New-build gentrification: its histories, trajectories, and critical geographies', *Population, Space and Place*, 16(5), pp. 395–411. Available at: https://doi.org/10.1002/PSP.584.

Davies, C. *et al.* (2009) 'User needs and implications for modelling vague named places', *Spatial Cognition and Computation*, 9(3), pp. 174–194. Available at: https://doi.org/10.1080/13875860903121830/ASSET/8023174A-65E1-42EA-8B88-F17C39527763/ASSETS/IMAGES/HSCC_A_412356_O_F0003G.GIF.

Debrunner, G. and Hartmann, T. (2020) 'Strategic use of land policy instruments for affordable housing – Coping with social challenges under scarce land conditions in Swiss cities', *Land Use Policy*, 99, p. 104993. Available at: https://doi.org/10.1016/J.LANDUSEPOL.2020.104993.

Dietrich-Gsenger, M. and Seethaler, J. (2019) 'Sozialdemografische Merkmale', in T. Hanitzsch, J. Seethaler, and V. Wyss (eds) *Journalismus in Deutschland, Österreich und der Schweiz* , pp. 51–70. Available at: http://www.springer.com/series/15233 (Accessed: 10 September 2024).

Eitle, E. (2001) 'Von den Klischees zu den Visionen ', *TagesAnzeiger*, 24 August.

Elliott-Cooper, A., Hubbard, P. and Lees, L. (2020) 'Moving beyond Marcuse: Gentrification, displacement and the violence of un-homing', *Progress in Human Geography*, 44(3), pp. 492–509. Available at: https://doi.org/10.1177/0309132519830511.

Feusi, A. and Hofer, K. (2020) 'Zürich Schwamendingen: Ein dörfliches Quartier im rasanten Wandel', *NZZ*, 5 August. Available at: https://www.nzz.ch/zuerich/zuerich-schwamendingen-ein-doerfliches-quartier-im-rasanten-wandel-ld.1568989 (Accessed: 13 August 2024).

Franklinturm (2024) *Franklinturm - Die Zukunft Ihres Business steht in Zürich Nord*. Available at: https://franklinturm.ch/ (Accessed: 2 August 2024).

Fritzsche, D. (2018) *Wie ein Deckel über der Autobahn das Quartier Schwamendingen nach 40 Jahren wieder vereinen soll*, *NZZ*. Available at: https://www.nzz.ch/zuerich/einhausung-schwamendingen-wie-das-quartier-zusammenwachsen-soll-ld.1442236 (Accessed: 7 September 2024).

FSO (2024) *Eidg. Gebäude- und Wohnungsregister (GWR)*. Available at: https://www.housing-stat.ch/de/madd/public.html (Accessed: 2 September 2024).

Gimes, M. (2015) ' Das Zürcher Quartier Schwamendingen ist eine landesweite Marke: Bünzliquartier für die einen, Problemviertel für die anderen – und von überall hört man die Autobahn. Jetzt soll sich

alles ändern. Die Fotografin Ruth Erdt hat die Entwicklung in Bildern festgehalten. ', *Das Magazin*, 18 April.

Hackworth, J. and Smith, N. (2001) 'The changing state of gentrification', *Tijdschrift voor Economische en Sociale Geografie*, 92(4), pp. 464–477. Available at: https://doi.org/10.1111/1467-9663.00172.

Hirschberg, J. and Manning, C.D. (2015) 'Advances in natural language processing', *Science*, 349(6245), pp. 261–266. Available at: https://doi.org/10.1126/SCIENCE.AAA8685.

Huber, M. (2012) 'Das teuerste Schulhaus von Zürich', *Tages-Anzeiger*, 18 September. Available at: https://www.tagesanzeiger.ch/das-teuerste-schulhaus-von-zuerich-902540366865 (Accessed: 16 August 2024).

Jain, S. *et al.* (2021) 'Nowcasting Gentrification Using Airbnb Data', *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1). Available at: https://doi.org/10.1145/3449112.

Jiang, N. *et al.* (2023) 'Leveraging newspapers to understand urban issues: A longitudinal analysis of urban shrinkage in Detroit', *Environment and Planning B: Urban Analytics and City Science* [Preprint]. Available at: https://doi.org/10.1177/23998083231204695.

Johnson, M.N.P. and McLean, E. (2020) 'Discourse Analysis', *International Encyclopedia of Human Geography, Second Edition*, pp. 377–383. Available at: https://doi.org/10.1016/B978-0-08-102295-5.10814-5.

Jones, K., Maynard, D. and Wartmann, F. (2022) 'Best Practice for Forensic Fishing: Combining Text Processing with an Environmental History View of Historic Travel Writing in Loch Lomond, Scotland', in R.S. Purves, O. Koblet, and B. Adams (eds) *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. Ubiquity Press, pp. 133–160. Available at: https://doi.org/10.5334/BCS.G.

Jungen, A. (2023) *Verdrängte Mieterin: «Dieses Ohnmachtsgefühl hallt nach»*, *SRF*. Available at: https://www.srf.ch/kultur/gesellschaft-religion/explodierende-mietpreise-verdraengte-mieterin-dieses-ohnmachtsgefuehl-hallt-nach (Accessed: 16 February 2024).

Kälin, A. (2022) 'Die Stadt Zürich tritt Land im Glattpark ab – zu einem Quadratmeterpreis wie am Paradeplatz', *NZZ*, 28 May. Available at: https://www.nzz.ch/zuerich/stadt-zuerich-tritt-land-im-glattpark-zu-paradeplatz-preisen-ab-ld.1686086 (Accessed: 13 August 2024).

Kaufmann, D. *et al.* (2023) *Erkenntnisse zum aktuellen Wohnungsnotstand: Bautätigkeit, Verdrängung und Akzeptanz*. Zürich. Available at: https://doi.org/10.3929/ethz-b-000603229.

Keller, E. and Bodmer, A. (2013) 'Wie Stadtplanung Geschichte macht', *WOZ*, 28 November. Available at: https://www.woz.ch/1348/neu-oerlikon/wie-stadtplanung-geschichte-macht (Accessed: 11 July 2024).

Koblet, O. and Purves, R.S. (2020) 'From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally', *Landscape and Urban Planning*, 197. Available at: https://doi.org/10.1016/j.landurbplan.2020.103757.

Kohler, A. *et al.* (2016) 'Erkunden Sie alle 2505 Zürcher Strassen ', *Neue Zürcher Zeitung*, 7 March. Available at: https://www.nzz.ch/zuerich/stadt-zuerich/in-den-strassen-von-zuerich-vom-aemmerliweg-zur-zwinglistrasse-ld.5453 (Accessed: 4 July 2024).

Kucera, A. (2011) 'Freitags Rückkehr zu den Wurzeln', *NZZ*, 17 September. Available at: https://www.nzz.ch/freitags_rueckkehr_zu_den_wurzeln-ld.666711 (Accessed: 4 September 2024).

von Ledebur, M. (2023) ‘Das geplante Zürcher Riesenschulhaus kostet pro Klasse 4 Millionen Franken. Wie kann das sein?as sein?’, *NZZ*, 1 June. Available at: https://www.nzz.ch/zuerich/das-geplante-zuercher-riesenschulhaus-kostet-pro-klasse-4-millionen-franken-wie-kann-das-sein-ld.1740229 (Accessed: 16 August 2024).

Ley, D. (1986) ‘Alternative Explanations for Inner-City Gentrification: A Canadian Assessment’, *Annals of the Association of American Geographers*, 76(4), pp. 521–535. Available at: https://doi.org/10.1111/j.1467-8306.1986.tb00134.x.

Loukaitou-Sideris, A., Gonzalez, S. and Ong, P. (2019) ‘Triangulating Neighborhood Knowledge to Understand Neighborhood Change: Methods to Study Gentrification’, *Journal of Planning Education and Research*, 39(2), pp. 227–242. Available at: https://doi.org/10.1177/0739456X1773089.

Luria, S. and Campos, R. (2022) ‘Greening a Post-Industrial City: Applying Keyword Extractor Methods to Monitor a Fast-Changing Environmental Narrative’, in R.S. Purves, O. Koblet, and B. Adams (eds) *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. London: Ubiquity Press, pp. 109–132. Available at: https://doi.org/10.5334/BCS.F.

Marcuse, P. (1985) ‘Gentrification, Abandonment, and Displacement: Connections, Causes, and Policy Responses in New York City’, *Washington University Journal of Urban and Contemporary Law*, 28(1), pp. 195–240. Available at: https://journals.library.wustl.edu/urbanlaw/article/id/8350/ (Accessed: 16 September 2024).

McKenzie, G. *et al.* (2018) ‘Identifying urban neighborhood names through user-contributed online property listings’, *International Journal of Geo-Information*, 7(10), p. 388. Available at: https://doi.org/https://doi.org/10.3390/ijgi7100388.

Murel, J. and Kavlakoglu, E. (2024) *What is Latent Dirichlet allocation | IBM*, *IBM*. Available at: https://www.ibm.com/topics/latent-dirichlet-allocation (Accessed: 20 August 2024).

Naismith, L. and Murphy, L. (2023) ‘Refashioning place and new-build gentrification: The material and symbolic redevelopment of Three Kings, Auckland’, *New Zealand Geographer*, 79(1), pp. 3–14. Available at: https://doi.org/10.1111/NZG.12349.

NOERD (2024) *NOERD: Gewerbehaus der Kreativen in Zürich Nord | NŒRD*. Available at: https://www.noerd.ch/ (Accessed: 18 July 2024).

NZZ. (1996) ‘Planung über alle Grenzen hinweg : Stadtbahn und «Grün 2005» für Leutschenbach und Opfikon’, *NZZ*, 12 June.

NZZ. (2000) ‘Schönes neues Oerlikon?’, *NZZ*, 24 February.

NZZ (2002) ‘Ungewisse Zukunft | NZZ’, *NZZ*, 17 September. Available at: https://www.nzz.ch/article8DZ9C-ld.225141 (Accessed: 13 August 2024).

NZZ (2008) ‘Wuchtige Bauten sorgen für günstige Wohnungen und Kritik’, 3 November. Available at: https://www.nzz.ch/wuchtige_bauten_sorgen_fuer_guenstige_wohnungen_und_kritik-ld.524069 (Accessed: 12 July 2024).

NZZ (2009) ‘Wohn-Kampfzone Seefeld’, *NZZ*, 6 December. Available at: https://www.nzz.ch/wohn-kampfzone_seefeld-ld.937933 (Accessed: 1 March 2024).

Olson, A.W. *et al.* (2021) 'Reading the city through its neighbourhoods: Deep text embeddings of Yelp reviews as a basis for determining similarity and change', *Cities*, 110. Available at: https://doi.org/10.1016/j.cities.2020.103045.

Olsson, J. (2021) *Modeling topics and semantic similarity with data from YouTube community discussions*. Uppsala Universitet. Available at: http://www.teknat.uu.se/student.

Pandur, M.B., Dobša, J. and Kronegger, L. (2020) 'Topic Modelling in Social Sciences: Case Study of Web of Science', in *Conference: Central European Conference on Intelligent and Information Systems*. Varazdin. Available at: https://www.researchgate.net/publication/344517050.

Pegler, C., Li, H. and Pojani, D. (2020) 'Gentrification in Australia's largest cities: a bird's-eye view', *Australian Planner*, 56(3), pp. 191–205. Available at: https://doi.org/10.1080/07293682.2020.1775666.

Phillips, M. *et al.* (2021) 'Re-placing displacement in gentrification studies: Temporality and multi-dimensionality in rural gentrification displacement', *Geoforum*, 118, pp. 66–82. Available at: https://doi.org/10.1016/J.GEOFORUM.2020.12.003.

Pont, M.B. *et al.* (2021) 'Systematic review and comparison of densification effects and planning motivations', *Buildings and Cities*, 2(1), pp. 378–401. Available at: https://doi.org/10.5334/BC.125.

Purcell, M. (2013) 'POSSIBLE WORLDS: HENRI LEFEBVRE AND THE RIGHT TO THE CITY', *Journal of Urban Affairs*, 36(1), pp. 141–154. Available at: https://doi.org/10.1111/juaf.12034.

Purves, R.S. *et al.* (2018) *Geographic Information Retrieval: Progress and Challenges in Spatial Search of Text*, *Foundations and Trends® in Information Retrieval*. now Publishers Inc. Available at: https://doi.org/10.1561/9781680834130.

Purves, R.S., Koblet, O. and Adams, B. (2022a) 'Analysing Environmental Narratives Computationally', in *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. London: Ubiquity Press, pp. 43–84. Available at: https://doi.org/10.5334/bcs.c.

Purves, R.S., Koblet, O. and Adams, B. (2022b) 'Introduction', in *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. London: Ubiquity Press, pp. 3–16. Available at: https://doi.org/10.5334/BCS.A.

Purves, R.S., Koblet, O. and Adams, B. (2022c) 'Surveying the Terrain and Looking Forward', in R.S. Purves, O. Koblet, and B. Adams (eds) *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. London: Ubiquity Press, pp. 233–244. Available at: https://doi.org/10.5334/bcs.l.

Rérat, P. *et al.* (2009) 'From urban wastelands to new-build gentrification: The case of swiss cities', *Population, Space and Place*, 16(5), pp. 429–442. Available at: https://doi.org/10.1002/psp.595.

Rey, C. (2022) 'Plötzlich ging es schnell: Zürich übertrifft den Bevölkerungsrekord von 1962 – das hat einen besonderen Grund', *NZZ*, 30 May. Available at: https://www.nzz.ch/zuerich/einwohner-stadt-zuerich-uebertrifft-bevoelkerungsrekord-von-1962-ld.1686437 (Accessed: 29 February 2024).

Rice, J.L. *et al.* (2020) 'Contradictions of the Climate-Friendly City: New Perspectives on Eco-Gentrification and Housing Justice', *International Journal of Urban and Regional Research*, 44(1), pp. 145–165. Available at: https://doi.org/10.1111/1468-2427.12740.

Rohrer, J. (2015) 'Seltsame Sprachkosmetik an der Zürcher Haltestelle', *Tages-Anzeiger*, 21 December. Available at: https://www.tagesanzeiger.ch/seltsame-sprachkosmetik-an-der-zuercher-haltestelle-283848116699 (Accessed: 15 August 2024).

Ruchirawat, N. (2018) *Collocations — identifying phrases that act like single words in Natural Language Processing*, *Medium*. Available at: https://medium.com/@nicharuch/collocations-identifying-phrases-that-act-like-individual-words-in-nlp-f58a93a2f84a (Accessed: 20 August 2024).

Sarkar, D. (2016) *Text Analytics with Python - A Practical Real-World Approach to Gaining Actionable Insights from Your Data*. 1st edn, *Text Analytics with Python*. 1st edn. Bangalore: Apress. Available at: https://doi.org/10.1007/978-1-4842-2388-8.

SBB (2024) *Zahlen und Fakten - Bahnhöfe*. Available at: https://reporting.sbb.ch/bahnhoefe?highlighted=ac0cc8146b1cce750ced69953d7a6f79&years=1,4,5,6,7&scroll=1498 (Accessed: 30 August 2024).

Scagliola, S. (2024) 'ISOS als Problem: ein lokales Thema', *P.S.*, 8 Autumn. Available at: https://www.pszeitung.ch/isos-als-problem-ein-lokales-thema/ (Accessed: 30 August 2024).

Schmid, C. (2012) 'Henri Lefebvre, the right to the city and the new metropolitan mainstream', in N. Brenner, P. Marcuse, and M. Mayer (eds) *Cities for people, not for profit : critical urban theory and the right to the city*. Routledge, pp. 42–62. Available at: https://doi.org/10.4324/9780203802182.

SGE (2024) *Neubausiedlung Obsthalden*. Available at: https://www.eigengrund.ch/liegenschaften/bauprojekte/obsthalden/erstvermietung/erstvermietung-obz-abgelaufen.html (Accessed: 2 August 2024).

Slater, T. (2011) 'Gentrification of the City', in *The New Blackwell Companion to the City*. Wiley, pp. 571–585. Available at: https://doi.org/10.1002/9781444395105.ch50.

Smith, N. (1987) 'Gentrification and the Rent Gap', *Annals of the Association of American Geographers*, 77(3), pp. 462–465.

SRF (2021) *Abstimmung Stadt Zürich - Mega-Schulhaus: «Teurer Palast» oder die «ideale Lösung»? - News - SRF*, *SRF*. Available at: https://www.srf.ch/news/abstimmungen-18-juni-2023/abstimmung-stadt-zuerich-mega-schulhaus-teurer-palast-oder-die-ideale-loesung (Accessed: 16 August 2024).

SRF (2023) *Schweiz aktuell vom 02.08.2023* . Available at: https://www.srf.ch/play/tv/schweiz-aktuell/video/schweiz-aktuell-vom-02-08-2023?urn=urn:srf:video:1da0949d-9bdb-46a3-a279-203672338efd (Accessed: 7 September 2024).

Stahel, T. (2006) *Wo-Wo-Wonige! Stadt-und wohnpolitische Bewegungen in Zürich nach 1968*. Universität Zürich. Available at: https://doi.org/https://doi.org/10.5167/uzh-163425.

Stopwords Iso (2020) *stopwords-de: German stopwords collection*, *GitHub*. Available at: https://github.com/stopwords-iso/stopwords-de/tree/master?tab=readme-ov-file (Accessed: 2 September 2024).

Sturzenegger, M. and Lutz, M. (2017) 'Wo die Seefeldisierung heute stattfindet', *Tages Anzeiger*, 28 February. Available at: https://www.tagesanzeiger.ch/wo-die-seefeldisierung-heute-stattfindet-610336283536 (Accessed: 1 March 2024).

Sturzenegger, S. (2022) *What Makes an Urban Public Space Popular? A Data-Based Analysis of Existing Urban Public Spaces in the City of Zurich*. Master Thesis. University of Zurich.

Swissdox@LiRI (2024) *Swissdox@LiRI*, *LiRI Wiki*. Available at: https://liri.linguistik.uzh.ch/wiki/langtech/swissdox/start#retrieveddatasets (Accessed: 16 February 2024).

Taylor, J.E. and Adams, B. (2022) 'The Wild Process: Constructing Multi-Scalar Environmental Narratives', in R.S. Purves, O. Koblet, and B. Adams (eds) *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. London: Ubiquity Press, pp. 161–178. Available at: https://doi.org/10.5334/BCS.H.

Teller, J. (2021) 'Regulating urban densification: what factors should be used?', *Buildings and Cities*, 2(1), pp. 302–317. Available at: https://doi.org/10.5334/BC.123.

Thalmann, P. (2012) 'Housing market equilibrium (almost) without vacancies', *Urban Studies*, 49(8), pp. 1640–1655. Available at: https://doi.org/10.1177/0042098011417902.

The World Bank (2023) *Urban Development*. Available at: https://www.worldbank.org/en/topic/urbandevelopment/overview.

Tolfo, G. and Doucet, B. (2021) 'Gentrification in the media: the eviction of critical class perspective', *Urban Geography*, pp. 1–22. Available at: https://doi.org/10.1080/02723638.2020.1785247.

Troxler, I. (2010) 'Neubauten sollen die Abwärtsspirale stoppen', *NZZ*, 6 January. Available at: https://www.nzz.ch/neubauten_sollen_die_abwaertsspirale_stoppen-ld.953606 (Accessed: 18 July 2024).

Troxler, I. (2013) 'Das erste von sieben neuen Schulhäusern', *NZZ*, 27 May. Available at: https://www.nzz.ch/zuerich/das-erste-von-sieben-neuen-schulhaeusern-ld.1053138 (Accessed: 16 August 2024).

Troxler, I. (2019) 'Bauen in den Städten: Statt Verdichtung zu verteufeln, sollten wir sie attraktiver gestalten', *NZZ*, 30 July. Available at: https://www.nzz.ch/meinung/hochhaeuser-als-chance-warum-dichte-zuerich-guttut-ld.1494466 (Accessed: 7 September 2024).

UZH (2024) *Rangkorrelation nach Spearman*, *Methodenberatung*. Available at: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/rangkorrelation.html (Accessed: 4 September 2024).

Waidmatt (2024) *Siedlung 16 Afaltra, Riedenhaldenstrasse und In Böden, Zürich*. Available at: https://www.waidmatt.ch/siedlungen/zuerich-affoltern/siedlung-16/ (Accessed: 2 August 2024).

Wartena, C. (2023) *The Hanover Tagger (Version 1.1.0) Lemmatization, Morphological Analysis and POS Tagging in Python*. Hannover.

Widmer, C.; and Kübler, D. (2014) *Regenerating urban neighbourhoods in Europe: eight case studies in six European countries*. Aarau. Available at: https://doi.org/10.5167/uzh-104988.

Wohnbaugenossenschaften Schweiz (2024) *Was ist eine Genossenschaft?* Available at: https://www.wbg-schweiz.ch/information/genossenschaftlich_wohnen/was_ist_eine_genossenschaft (Accessed: 30 August 2024).

Zenebe, A., Brown-Robertson, L. and Mayo, K. (2018) 'Discovery of insights on gentrification using analytics from Twitter', in J.D. Belloit, C.J. Myers, and N.C. Sigmond (eds) *Northeastern Association of Business, Economics and Technology Proceedings*. Pennsylvania, pp. 368–377.

Appendix

In the appendix, the nodes and figures for the collocation results as well as the topic modelling that are described in the results part but not shown can be found.

Appendix A: Nodes for the Collocation Analysis

Table 24: List of words investigated in the collocation analysis, their translation and if results were found.

Keyword	Translation	Results?	Keyword	Translation	Results?
abreissen	demolish	no	mehrfamilienhaus	Apartment building	yes
abriss	demolition	no	mieten	Rent	yes
affoltern	Affoltern (district)	yes	neubau	New build	yes
areal	area	yes	oerlikon	Oerlikon (district)	yes
arealentwicklung	Areal development	yes	quartier	District	yes
aufschwung	upswing	yes	richtplan	Structure plan	yes
aufsteigend	Ascending (in terms of upswinging neigbhourhood)	yes	saatlen	Saatlen (neighbourhood)	yes
aufwerten	upgrade	yes	sanieren	Renovate	no
aussenentwicklung	Greenfield development	no	sanierung	Renovation	yes
auzelg	Auzelg (neigbhourhood)	yes	schwamendingen	Schwamendingen (district)	yes
bauen	building	yes	seebach	Seebach (district)	yes
brache	Fallow	yes	siedlung	settlement	no
einfamilienhaus	Single-family house	yes	sondernutzungsplan	Special land use plan	yes
einhausung	Enclosure (project to enclose the highway passing through Schwamendingen)	yes	teuerung	Price increase	yes
expat	Expat	no	verdichten	densify	yes
gartenstadt	Garden city (city	yes	verdichtung	densification	yes
genossenschaft	Cooperative	yes	verdrängen	Displace	yes
gentrifizieren	Gentrify	yes	verdrängung	Displacement	yes
hipster	Hipster	yes	verteuern	increasing price	yes
immobilie	Real estate	yes	verteuerung	Increase of price	yes
industriebrache	Industrial fallow	yes	wohnen	Living	yes
innenentwicklung	Brownfield development	yes	wohnung	Apartment	yes
instandsetzen	Restore	yes	wohnungsnot	Housing shortage	yes
instandsetzung	Restoration	yes	yuppies	yuppie	no
investieren	Invest	yes	zuzug	Moving in	yes
investoren	Investor	no	überbauen	overbuild	yes
kreativ	Creative	yes	überbauung	Overbuilding	yes
lärm	noise	yes			

Appendix B: Collocation Word Clouds

Collocation of neighbourhood names

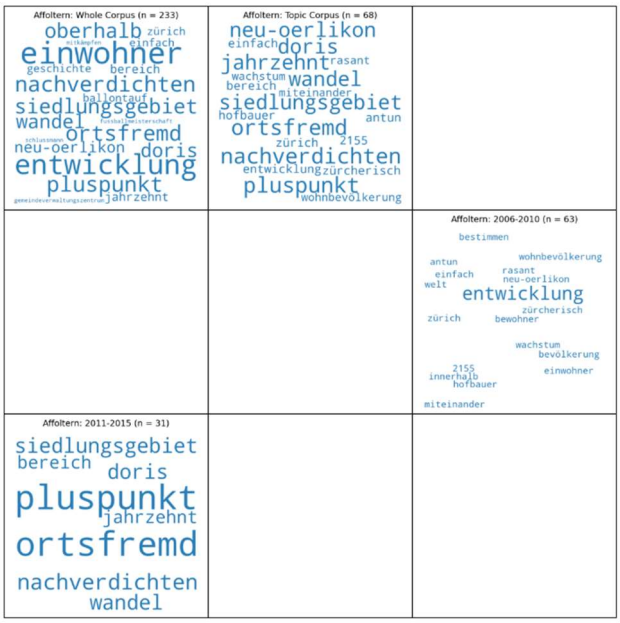

Figure 19: Collocation results "Affoltern" for different sub-corpora. Most collocate relate to urban development and population growth.

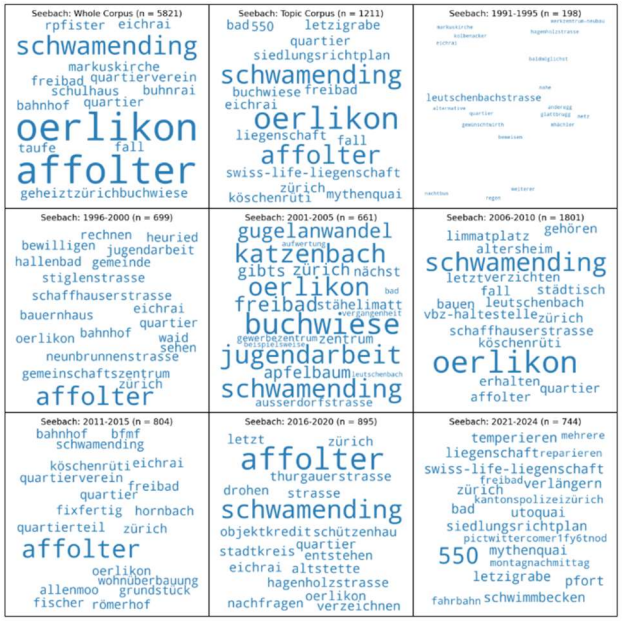

Figure 20: Collocation results "Seebach" for different sub-corpora. Seebach collocates with a variety topics and locations.

Figure 21: Collocation results for "Oerlikon" for different sub-corpora. Many collocates are about point of interests and the machine factory Bührle.

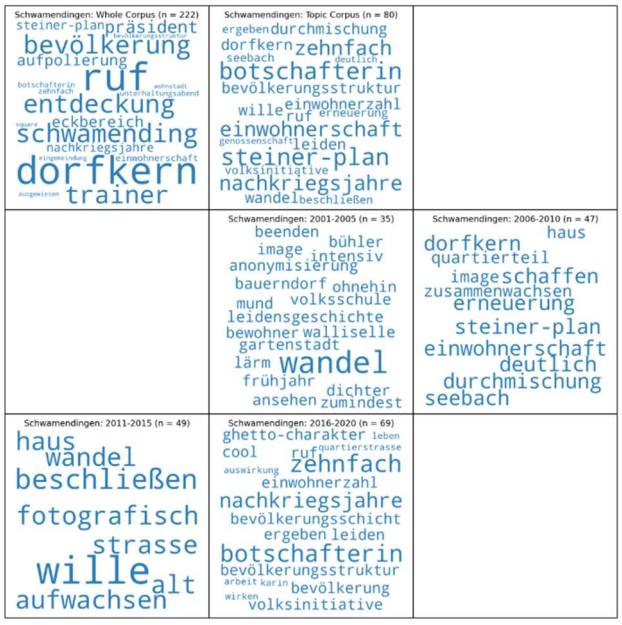

Figure 22: Collocation results for "Schwamendingen" for different sub-corpora. Many collocates relate to the history, image or population growth.

Figure 23: Collocation results for "Leutschenbach" for different sub-corpora. A change on discourse can be seen as Leutschenbach becomes known for the neighbourhood and not only the TV-studios.

Figure 24: Collocation results for "Mattenhof" for different sub-corpora. Many places outside of Zurich such as Breitenrain and Kriens can be found, showing that Mattenhof is an ambiguous location.

Figure 25: Collocation results for "Neu-Oerlikon" for different sub-corpora. The results show that Neu-Oerlikon is a development area that becomes only known after the transformation has started in 2001.

Collocation of the five keywords

Figure 26: Collocation results for "Gentrifizierung" (gentrification) for different sub-corpora. The term becomes better known after 2006.

Figure 27: Collocation results for "Aufwertung" (uplifting) for different sub-corpora. It shows that uplifting is used in diverse contexts from nature to public spaces and neighbourhoods.

Figure 28: Collocation results for "Aufwerten" (uplift) for different sub-corpora. The results are similar than for Aufwertung showing that they are interchangebly.

Figure 29: Collocation results for "Neubau" (new build) for different sub-corpora. Neubau is exclusively used in context of building and is used consistently over time.

Figure 30: Collocation results for "Sanierung" (renovation) for different sub-corpora. The use of the term has not changed much over the past 30 years and often relates to construction.

Figure 31: Collocation results for "Sanieren" (renovate) for different sub-corpora. The results are very similar than for Sanierung which shows that the terms are used interchangeably.

Figure 32: Collocation results for "Verdrängung" (displacement) for different sub-corpora. Verdrängung is mostly used regarding urban processes and since 2016 also in terms of gentrification and densification.

Figure 33: Collocation results for "Verdrängen" (displace) for different sub-corpora. The results are very diverse and refer to the displacement of different objects and subjects.

Collocations of other keywords of interest

Figure 34: Collocation results for "Industriebrache" (industrial fallow) for different sub-corpora. The term is often used in relation to conversion of certain areas to something new.

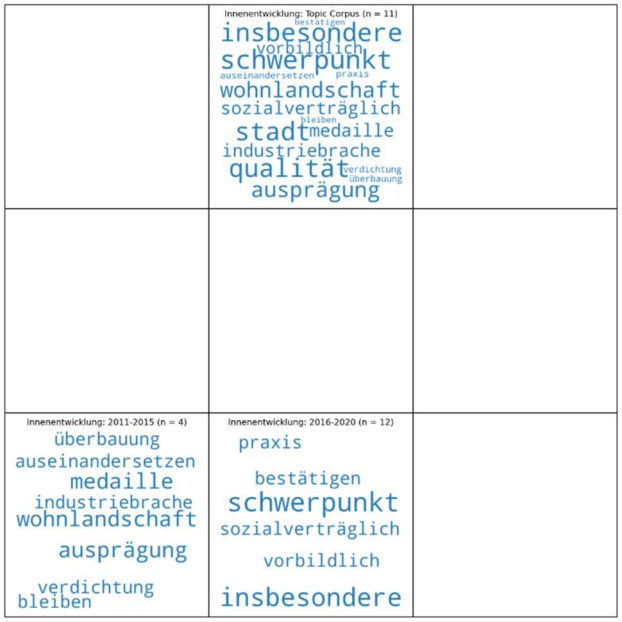

Figure 35: Collocation results for "Innenentwicklung" (inward development) for different sub-corpora. The term is only used after 2011 as from 2014 onward it is mandated by the spatial planning act.

Figure 36: Collocation results for "Quartier" (neighbourhood) for different sub-corpora. Neighbourhoods are described similarly over time as lively and mixed. After 2006 they are also characterised as growing.

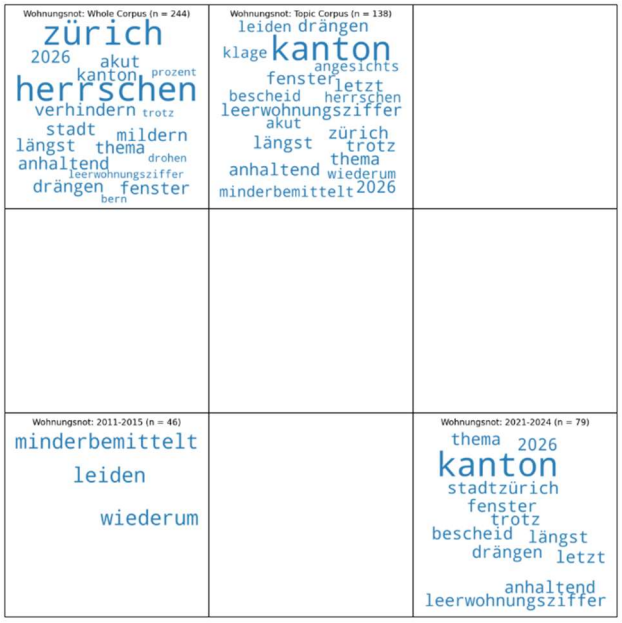

Figure 37: Collocation results for "Wohnungsnot" (housing shortage) for different sub-corpora. There are only two results for the period 2011 and 2021 where the housing shortage is described as lasting and persistent.

Appendix C: Topic Modelling Outputs

Topic Model Output for the Gentrification Sub-Corpus

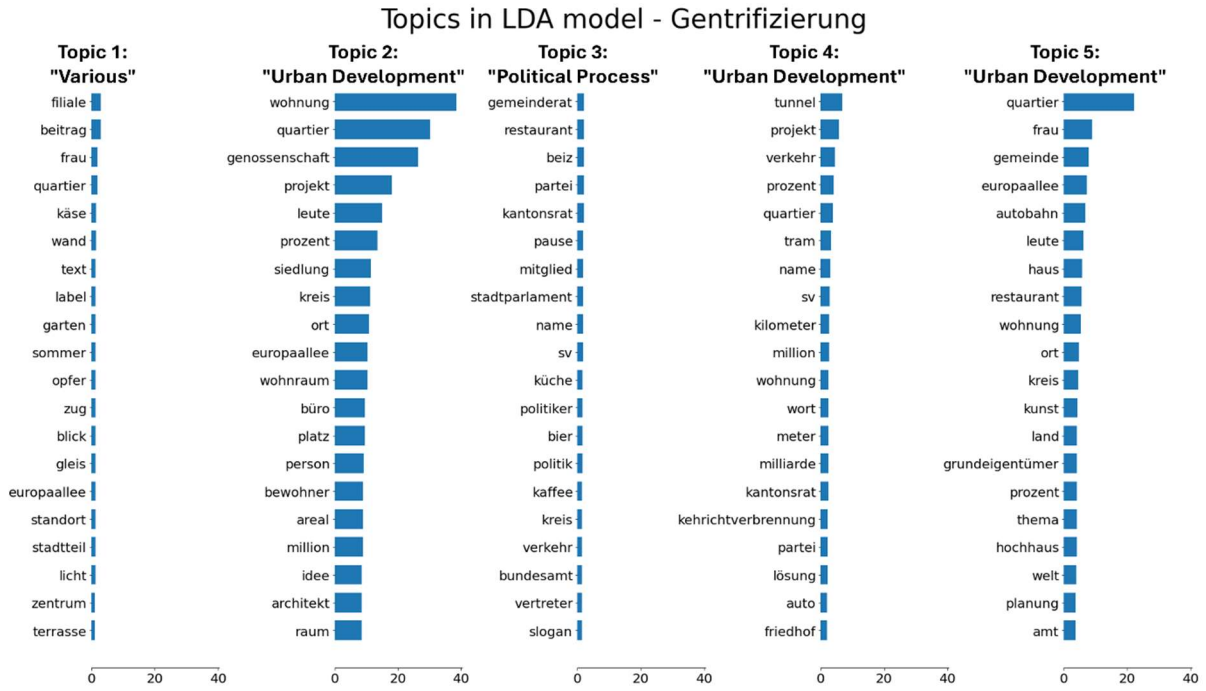

Figure 38: Topic model for the sub-corpus gentrification. Three of the five generated topics can be clearly distinguished whereby the urban development category is the most significant.

Topic Model Output for the New Build Sub-Corpus

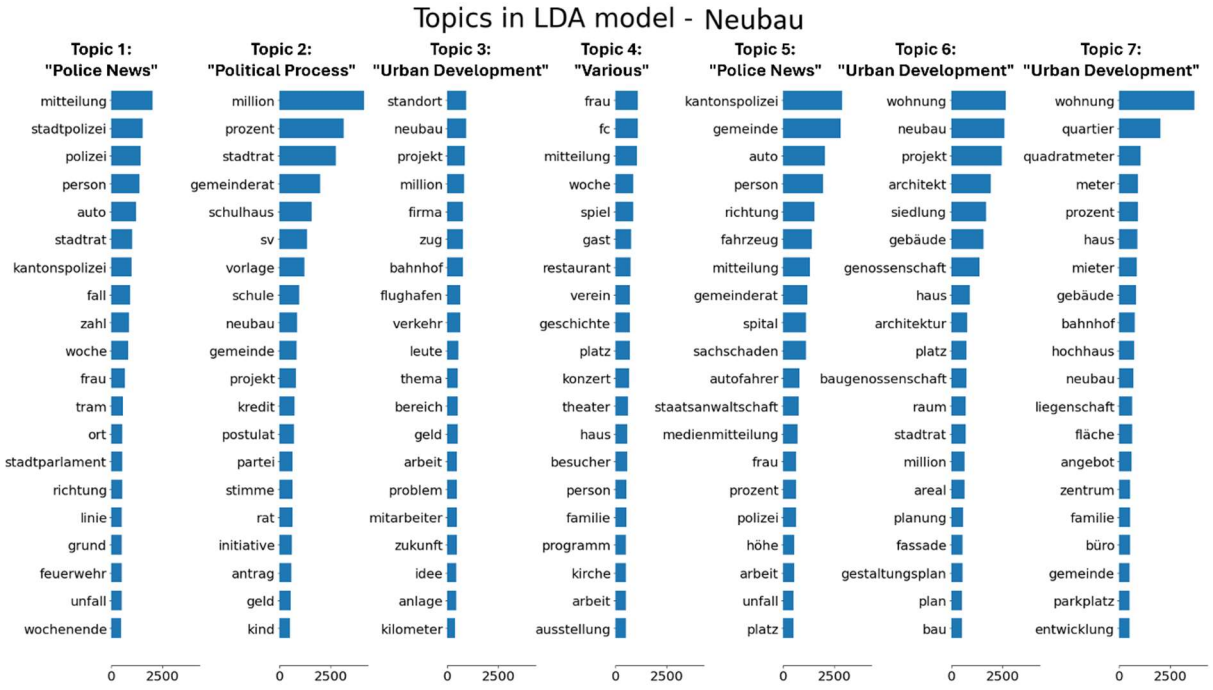

Figure 39: Topic model for the sub-corpus new build. Out of the seven generated topics, six can be clearly distinguished. However, some of them overlap thematically.

Topic Model Output for the Renovation Sub-Corpus

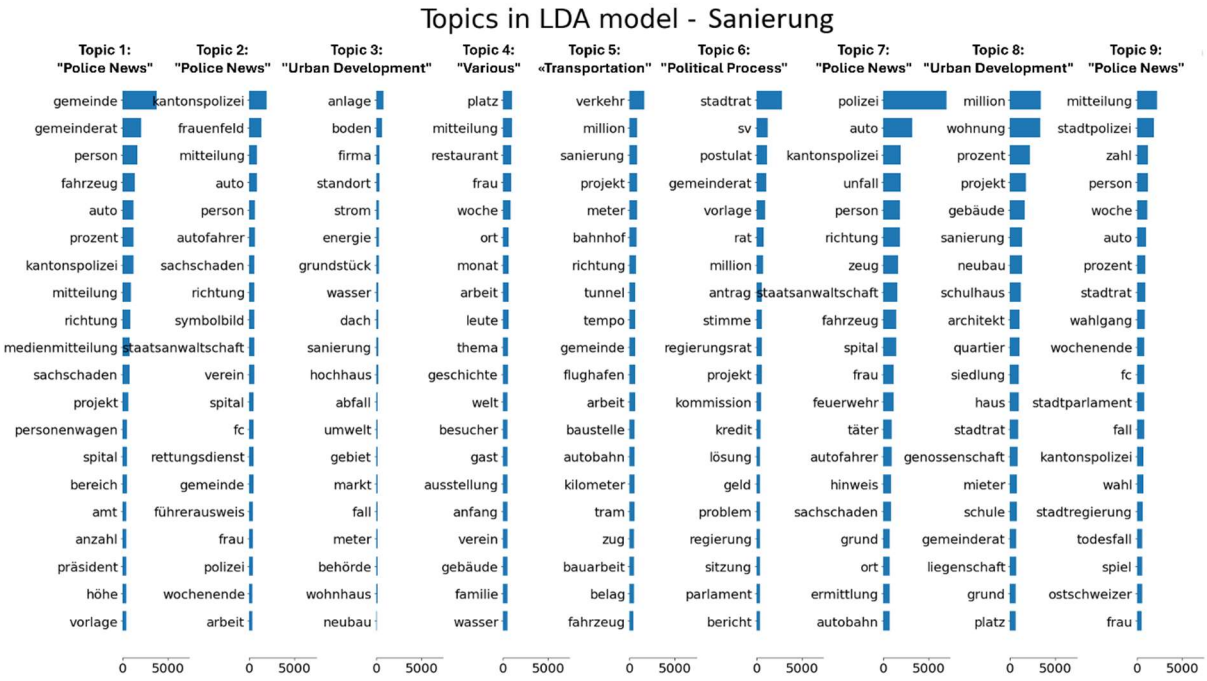

Figure 40: Topic model for the sub-corpus renovation. For the renovation sub-corpus, nine different topics were generated whereby a few of them are overlapping. All but one can be assigned a specific category.

Topic Model Output for the Displacement Sub-Corpus

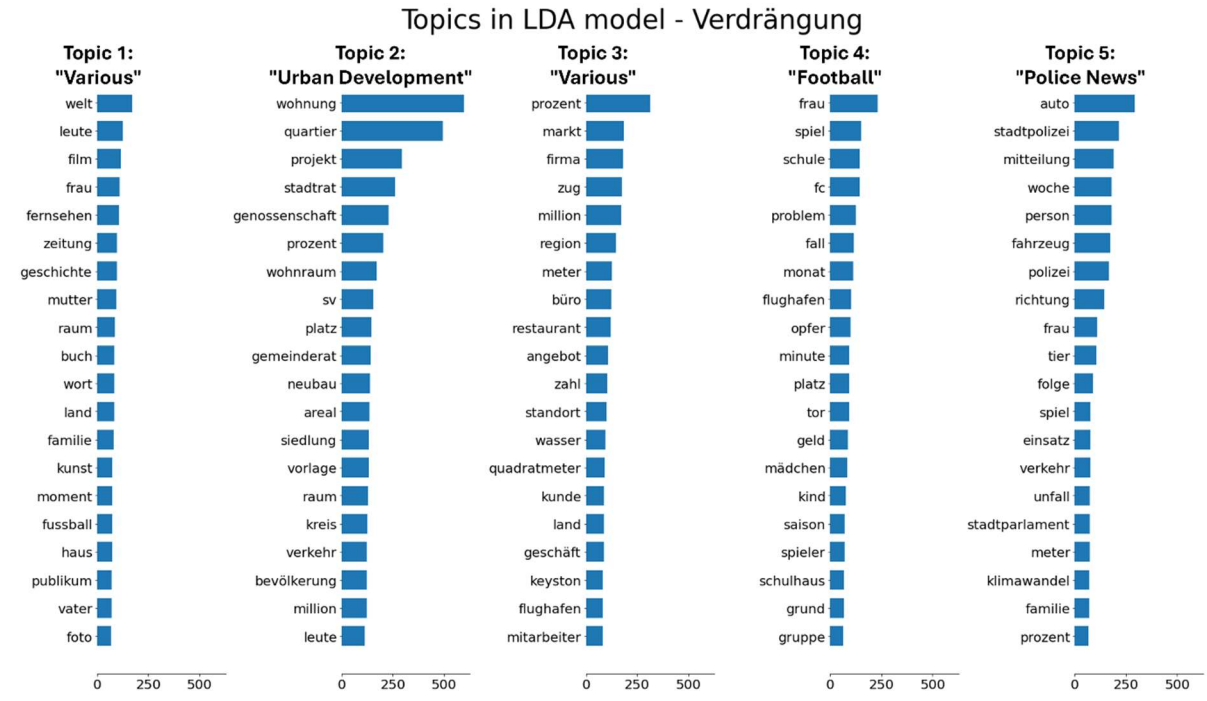

Figure 41: Topic model for the sub-corpus displacement. Two out of the five generated topics cannot be assigned to a broader category. Noticeable is also the category "football" which is only distinguishable in this sub-corpus.

Appendix D: Example Co-Occurrence Map

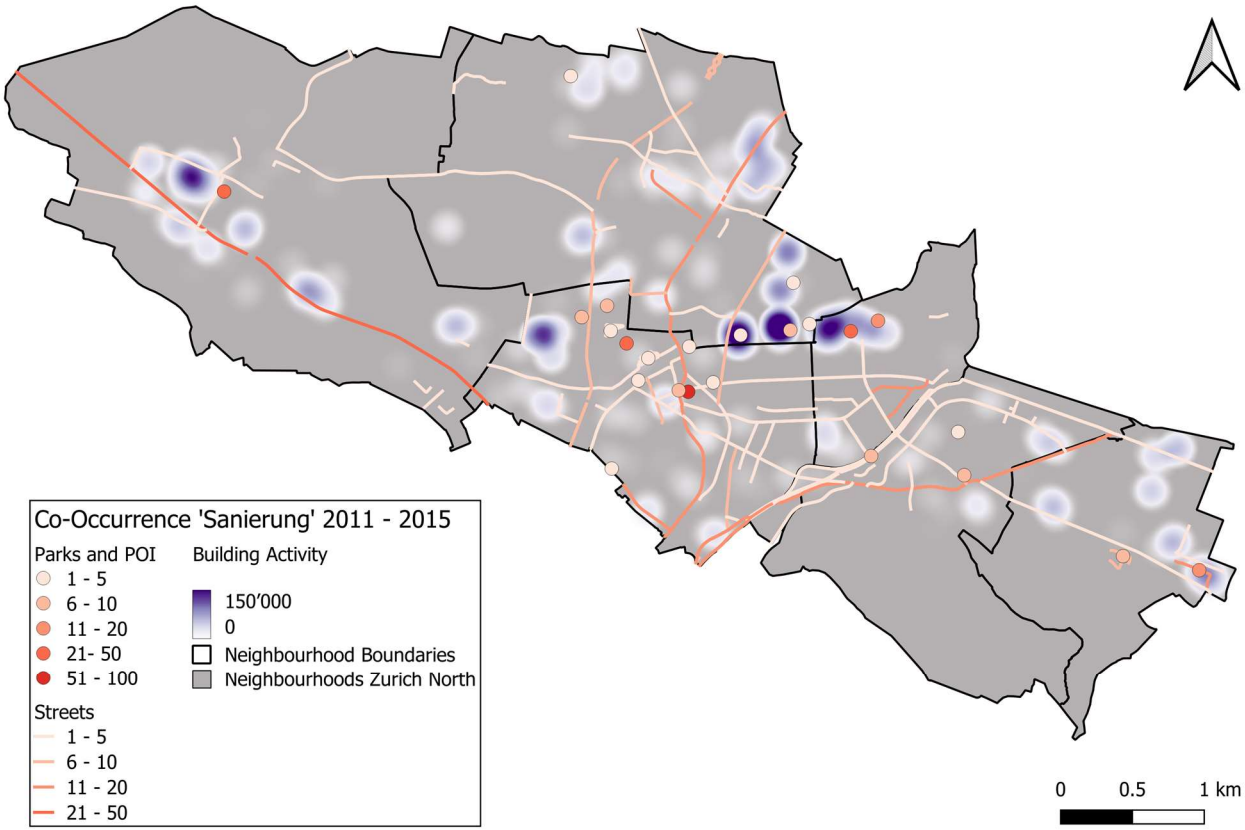

Figure 42: Co-Occurrence results for "Sanierung" (renovation) in the period 2011 - 2015. This is an enlargement of the map discussed in chapter 4.1.7 and serves as an example of the co-occurrence's maps in general. Depicted are the streets and points of interest. The deeper the red, the more frequent they appear in the corpus. The building activity is shown in purple per volume [m3].

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis. Furthermore, artificial intelligence in form of ChatGPT was used to generate and improve parts of the python code used in this thesis. For this publication, use was made of media data made available via Swissdox@LiRI by the Linguistic Research Infrastructure of the University of Zurich (see https://www.liri.uzh.ch/en/services/swissdox.html for more information).

Lynn Bolliger, 23.09.2024