

Failing to use the best words

Linn Seime Thistel, Statistics Norway, linn.thistel@ssb.no

Frode Larsen, Statistics Norway, frode.larsen@ssb.no

Abstract

This year, a new Language Usage Act was introduced in Norway. For the first time, the law clearly states that the public sector must write in a plain and correct manner, suitable for the target users. For Statistics Norway, the law formalizes and strengthens what has been a priority for many years. An important part of our mission is to write in order for the public to understand and use the figures we gather and interpret.

When we write about a complicated subject, our choice of words can matter a great deal to the reader, with respect to comprehensibility of the content and in order for the user to discover the text in the first place. The best words are not necessarily the most accurate ones, rather the words that best match the users' natural discourse might prove expedient.

In this paper, we will look at three different categories where we aim at using the best words for the public, but regrettably without succeeding. We will look at the reasons for why our efforts were insufficient, show examples, try to determine why we did not get the results we aimed for, and for each category point to a better solution for the future.

The three categories we will examine are: Search terms, simplifying categories and overcomplicated words.

We will show how we work systematically to analyze the users, and how we sometimes discover that they have evolved ahead of us linguistically.

We will show that efforts to help the user by simplifying words and categories sometimes can turn out wrong.

We will present the challenge of lingo or complicated language that can create a gap between us and the users.

Keywords: Dissemination, words, users

Content

Abstract	1
1. Introduction.....	3
1.1 <i>Clear language and the laws</i>	3
1.2 <i>Terms and definitions</i>	4
1.3 <i>Why it is our job to change</i>	4
2. Search terms – finding the balance between what people use and keeping continuity	5
2.1 <i>Getting to know the user</i>	6
3. Simplifying categories - Finding the balance between what is accurate and what is understandable.	7
3.1 <i>The misleading infographic</i>	9
3.2 <i>The first mistake: The number of cabins being built</i>	10
3.3 <i>The second mistake: The sizes of the cabins</i>	10
3.5 <i>So, what is the solution?</i>	10
3.4 <i>The balance between what is accurate and what is understandable</i>	12
4. Overcomplicated words – Finding the balance between smooth internal communications and inviting external communications	13
4.1 <i>So, what is the solution?</i>	14
5. Conclusion	14
6. Citations.....	15
7. Illustrations.....	15

1. Introduction

When we write about a complicated subject, our choice of words can matter a great deal to the reader, with respect to comprehensibility of the content, and in order for the user to discover the text in the first place. The best words are not necessarily the most accurate ones, rather the words that best match the users' natural discourse might prove expedient.

In this paper, we will look at the words we, as a statistical office, use when communicating to the user. In three chapters we will look at the balance between our and the users' perception. We call this the sweet spot of statistical understanding.

Reflecting on how things have been done wrong and how we can improve will be a common thread through the paper.

Chapter 1 – “Search terms” will explain how we work to find the words that meet and greet the user when they're searching. Chapter 2 – “simplifying categories” will go deep into one case. We will tell the story of one incident, and how trying to make improvements on the visualization ended up giving the wrong numbers. This is the paper's longest chapter. Chapter 3 – “overcomplicated words” will take a quick look at the how and why we in Statistics Norway use lingo unfamiliar to the public.

1.1 Clear language and the laws

The Norwegian Statistical Act (lovdata, 2022, 1) states that official statistics is the nation's mutual basis for facts, and important for a thriving democracy. For the statistics to be used as a means to democracy, the statistics need to be understood, both by the “experts” and the general public.

In 2022, a new Language Usage Act (lovdata, 2022, 2) was introduced in Norway. For the first time, the law clearly states that the public sector must write in a plain and correct manner, suitable for the target users.

For Statistics Norway, the law formalizes and strengthens what has been a priority for many years. An important part of our mission is to write in order for the public to understand and use the figures we gather and interpret.

To improve our clear language, Statistics Norway is now halfway through a project to increase awareness and competence in clear language: “The project clear language

2021-2023”. The project group has made a digital course, does regular marketing activities and holds workshops for other employees, to mention some activities. A new forum has recently been created, consisting of motivated “language contacts” from different departments in Statistics Norway. After 2023, the project will transform into another organizational form, but the focus will remain on improving Statistics Norway’s clear language.

1.2 Terms and definitions

In this paper, we have chosen to narrow the user group down to the general public. This is one of our, and the largest, defined target groups, and consists of people who use our information for personal use. They have a varying knowledge of statistics, and most often find our information through Google. For the sake of simplicity, we will refer to the general public as the user.

1.3 Why it is our job to change

Professional statistics and the users have different perceptions in regards to statistics. This can be illustrated by a Venn diagram. There is a sweet spot in the middle, and this can be small or large.

Improving the public’s statistical literacy is important to Statistics Norway. We hold free courses, make teaching materials for school children, and write explanatory texts on our web site, just to mention some initiatives. These initiatives pull the users in our direction, making the sweet spot larger, thus making it easier for Statistics Norway to use what we regard as the correct and accurate terms – The possibility that

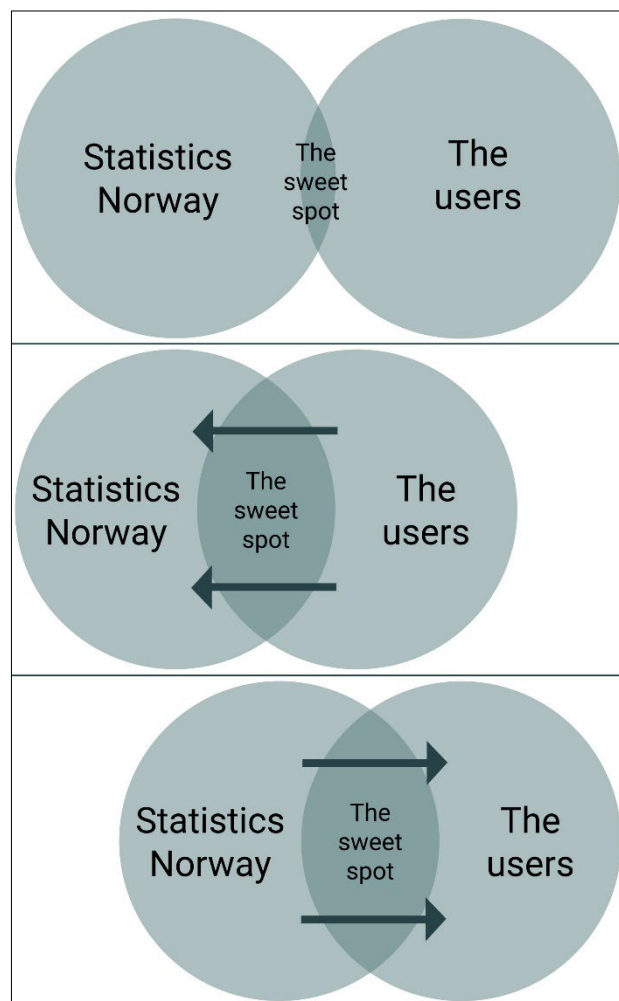


Figure 1: Venn diagrams showing the sweet spot between Statistics Norway and the users

these users understand is higher than that those with a lower statistical literacy will understand.

But the vast majority of the public will not be touched by our measures to teach. To reach those people, we need to make sure that *we* are moving closer to *them*, instead of them towards us. We do this by explaining and using words they understand. This will also make the sweet spot larger, but instead of projects and initiatives, this is a job that needs to seep through everything we do.

We are not saying we should stop working to improve the users' statistical literacy. What we are saying is the most effective way to make the sweet spot larger, is for us to use a language the user already is familiar with.

2. Search terms – finding the balance between what people use and keeping continuity

In 2013, 33% of the traffic on our web page, *ssb.no*, arrived through the front page. As of 2022, that number is down to 12%. The front page is no longer an important starting point, where we can guide the users to the right place. An easy take away from this is that the users have a question, but they don't necessarily know that they can find the answer on our web page. We need to guide them, and where we need to guide them *from*, is search engines. As the most common search engine is Google, we will hereafter only talk about Google.

Google is rapidly improving their algorithms, making it easier for the user to find relevant and good results based on their search words. But we still need to make it easy for Google to understand when our content is the right one, and make the user choose our page based on the preview in the results list. Choosing the right words is an important part of this.

We will in this chapter look at how to know which words and terms the users actually use, and what we can do to make our words match those. The chapter will end by taking a look at why it's not that simple.

2.1 Getting to know the user

Often, the first step of finding words that we use that differ from what the public uses, is as simple as noticing them ourselves. If we react, the users might also react. For example, up until recently, we used the term “beltemotorsykkel” (half-track motorcycle) when referring to “snøscooter” (snow mobiles). The first is an old term that is unquestionably correct for categorizing those types of vehicles, but the latter is what is commonly used for describing that specific vehicle today.

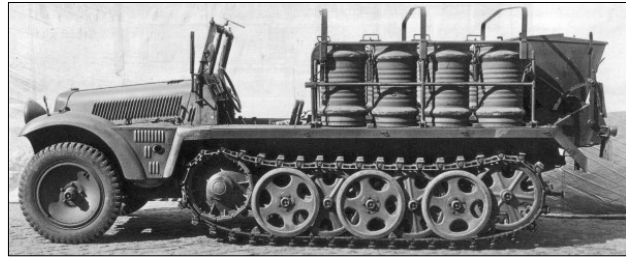


Illustration 1: “Beltemotorsykkel”/half-track-motorcycle



Illustration 2: “Snøscooter”/snow mobile

The next step is to investigate whether users in fact use “snøscooter” instead of

“beltemotorsykkel”. This can be done by looking at Google Trends. Google Trends shows how many times a specific word or phrase has been googled in a given period.

It is also possible to compare several search terms, like we have done in figure 2. It shows that “Snøscooter” (snow mobile) is the most googled word of the three we chose to compare. The second most googled is a different spelling of snow mobile. The yellow line, barely visible, shows the searches for the word “beltemotorsykkel”.

In this case, it is clear that we need to use the word snøscooter if we want the user to find our statistics about the theme.

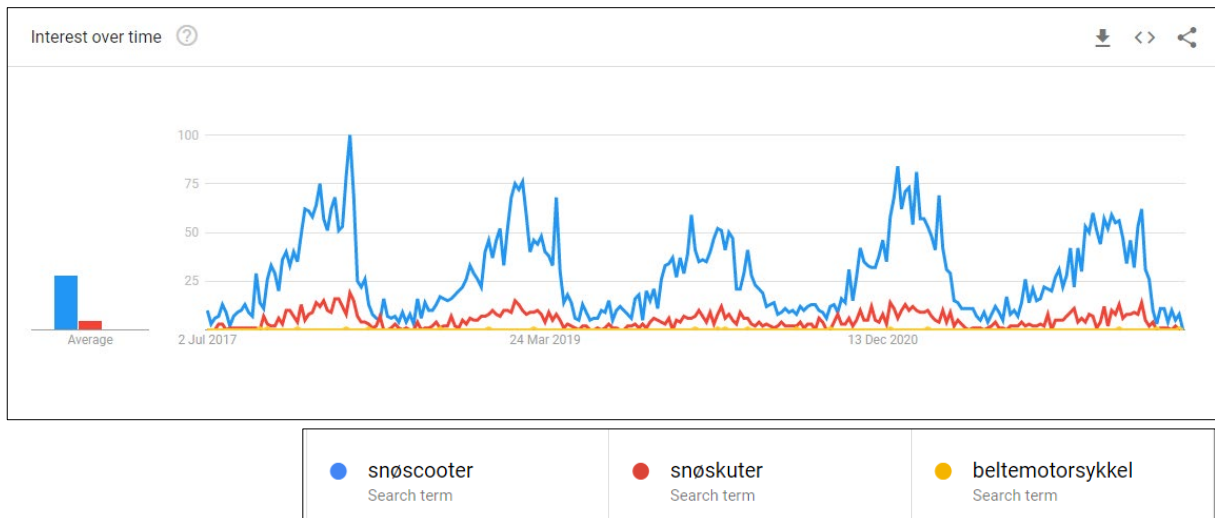


Figure 2: Google trends for the words "beltemotorsykkel", "snøscooter" and "snøskuter"

2.2 Why the terms are what they are

We need to be aware that “just changing” words and terms is not that simple. Often, there are good reasons for the term to be what it is, for example historical or international classification systems. In those cases we, as responsible for dissemination, need to find a balance where the commonly used term is the low threshold entry, and the correct term is used for the in depth statistics. There also needs to be a good bridge between the simple and the correct term.

3. Simplifying categories - Finding the balance between what is accurate and what is understandable.

In this chapter we will describe the case of the Norwegian cabins.

Let us start by defining the word cabin. Store Norske Leksikon define “hytte” (cabin) as «a Norwegian term, traditionally used about a small building, [...] used for short stay and with few or no modern conveniences. Later, the word has gotten the broader meaning



Illustration 3: Hytte/cabin. Small, traditional

holiday home or holiday house, with a large degree of comfort. The meaning can therefore also include huge buildings, almost like villas. (snl, 2022, my translation)

Just by the definition, we can see that there is a historical difference in what was and what is defined as a cabin, but the core of the word is the separate building.

In this chapter we must then separate between three important terms:

- Cabin («hytte»)
- Holiday home («fritidsbolig»)
- Holiday building («fritidsbygning»)

Our statistics are for holiday buildings, but as this is not a word commonly used in the public, both cabin and holiday home are easier for dissemination. But the meanings are not the same. As previously stated, the two first words describe a single home. Holiday building, on the other hand, means a whole building, regardless of how many homes are inside the building.

In recent years, a vast number of *apartment complexes* have been built. We do not have the exact numbers, so the best way to get an idea is to travel through Hemsedal, Geilo, Norefjell or another mountain destination. That is, of course, not enough to base any numbers on, but the view over apartment complexes will show that there definitely are *some*.



Illustration 4: Hytte/cabin. Large, modern



Illustration 5: Apartment complex



Illustration 6: Annex

To add a last cabin related term, *extensions* and *annexes* are built as part of or next to existing buildings. These are smaller buildings, and are not used for holiday homes on their own.

3.1 The misleading infographic

A little while ago, the communications department wanted to show that the size of the cabins built were increasing. This is the biggest change in the cabin market the last fifty years. We also wanted to illustrate that the number of cabins being built are increasing.

In infographics, there is rarely room for long explanations and reservations.

The result was the easy to understand caption “hytte” (cabin) instead

of a complex explanation of

how the number included cabins, other holiday homes, annexes, extensions and apartment buildings.

This illustration shows a footnote that was not there in the beginning. It was added after a while, to correct the misleading caption. After yet a while, it was clear the footnote still did not make the content of the infographic clear, and the infographic was taken down.

The infographic ended up not just being open for misinterpretation, but straight out wrong. In not just one, but in two ways.

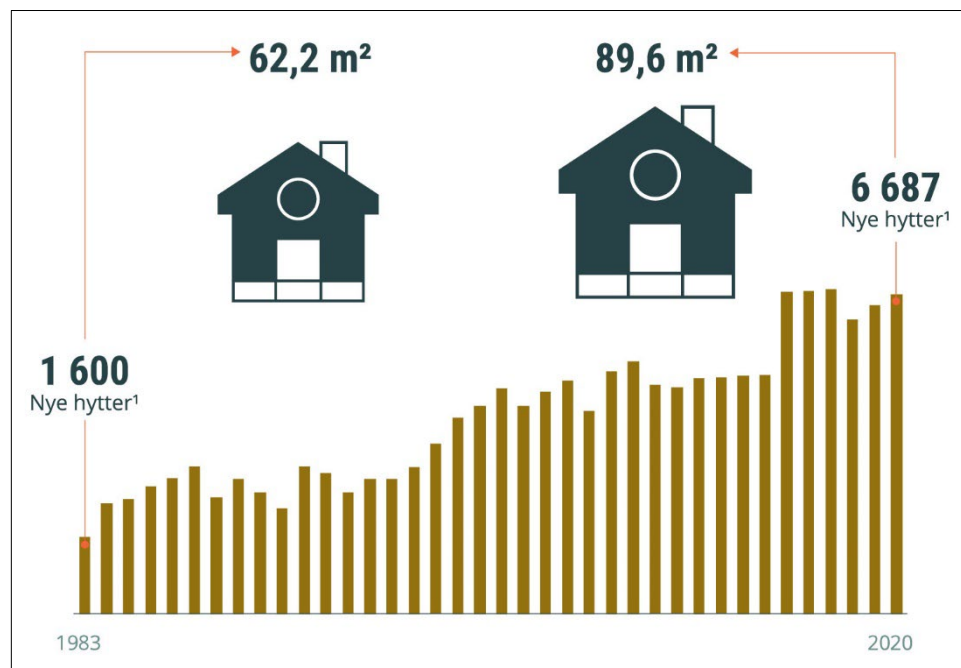


Figure 3: Infographic about cabins

3.2 The first mistake: The number of cabins being built

The graph gives a very specific number of cabins built in 2020: 6.687. Many users will interpret this to mean only cabins, but somebody might also include all sorts of holiday homes in their mental image. But in fact, this number shows *holiday buildings*, a category that includes cabins, extensions, annexes and apartment buildings. New cabins make up approximately 4500 cabins, and the rest is annexes, extensions and apartment buildings. The extensions and annexes will make the number seem larger than it is in reality (as these types of buildings are not for living in and should not be included). The apartment buildings will do the opposite, and count for *one* regardless of whether there are four or forty apartments in the building. This will make the number lower than it is in reality.

Giving this very complex number the title “nye hytter” (new cabins), is a failed attempt at simplification. Making it simpler actually makes it wrong. At the time, we did not know the number of cabins built, but it would be a mere coincidence if the number was 6.687.

3.3 The second mistake: The sizes of the cabins

The infographic shows that new cabins built was increasing in size over the years. Especially in 2015 the cabin sizes made a big jump. The average sizes in the infographic are not, in fact, only the sizes of cabins. Just as shown in chapter 3.2, this number hold not only the cabins, but also extensions and annexes (shrinking the average size) and apartment complexes (making the average size larger).

Could, for example, the sudden spike in 2015 be caused by several 2000 m² large apartment buildings? We didn't have the answer to that question, as we only had the numbers for cabins, extensions, annexes and apartment buildings in total. In chapter 3.5, we will show the correct numbers.

3.5 So, what is the solution?

This topic was discussed back and forth between the statistics divisions and the communications department, resulting in, first, a footnote and then the infographic being taken down. The problem was not mainly the faulty numbers, but

oversimplification of the titles that made it look like we knew something we did not know.

The case then led to the creation of brand-new statistics for cabins, this time with the intention of finding a good number for what the communications department had wanted to present in the infographic. The new statistics (ssb.no, 2022) were released in June 2022, and Statistics Norway are now able to say something about the Norwegian cabins.

The two graphs show the numbers from the infographic and the new numbers. The differences between the number sets are clear. We build fewer and larger holiday homes than we earlier presented. The trend is still somewhat the same between the two sets.

For these new numbers the label “feriebolig” (holiday home) consists of cabins and apartments used for leisure. Excluded from the numbers are extensions and annexes. Apartment complexes are no longer counted, only the apartments within. Notice that this means we still do not have the numbers for only cabins – the very thing we wanted to show in the infographic.

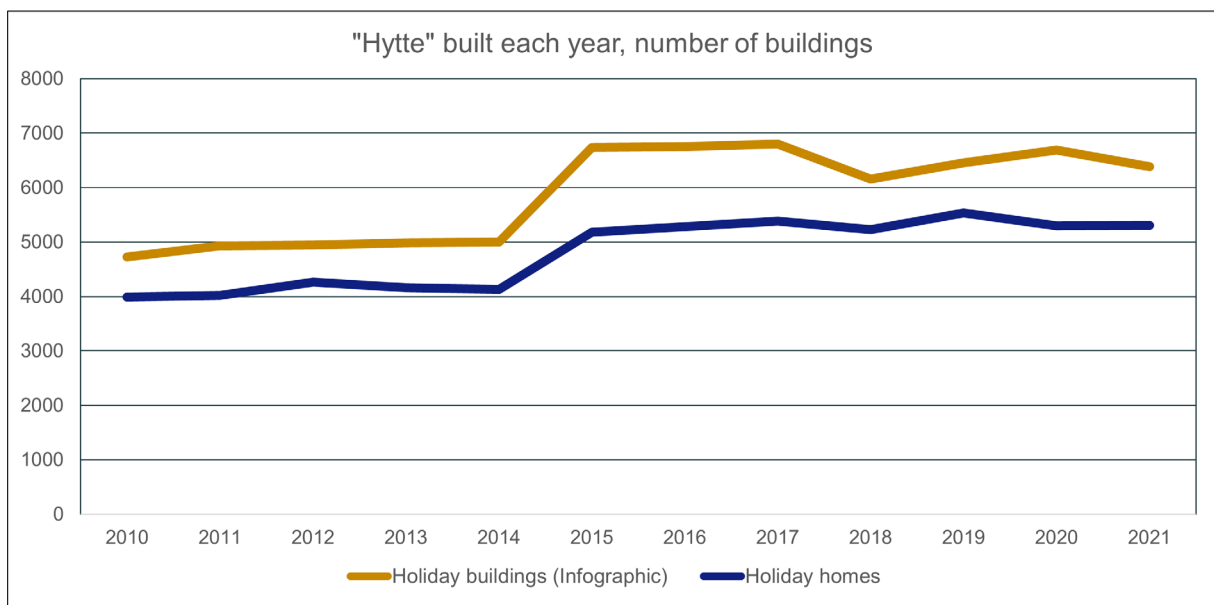


Figure 4: The number of holiday buildings and holiday homes built each year

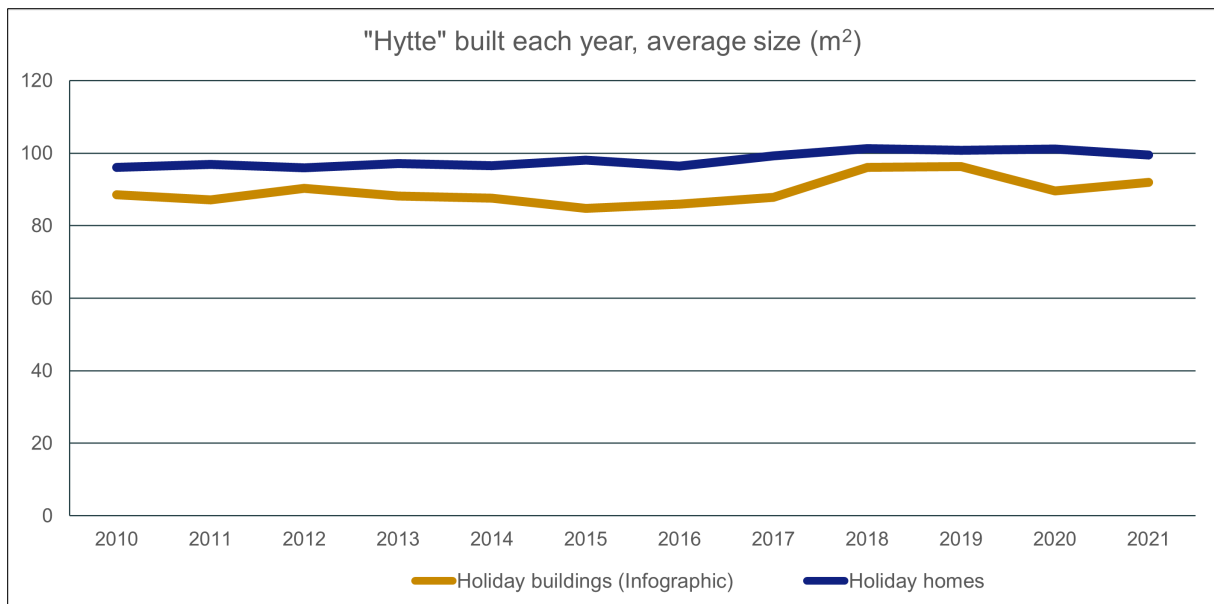


Figure 5: The sizes of holiday buildings and holiday homes built each year

Other statistics (Mathiesen, 2022) show that, on average, cabins are larger in size than apartments. This means that we can safely assume the numbers should be yet higher than the blue line in figure 5 shows.

Unfortunately, it is not possible to apply this new method to the earliest numbers, so the starting year for this series is 2010. This means it is not possible to say for sure the number and size of cabins earlier than 2010. Apartment complexes for leisure is a fairly new concept, and we can therefore assume the earlier numbers consist only of holiday homes and extensions/annexes. As the latter *decreases the average size and increases the number of buildings built*, we can assume the holiday homes built in the 1980s and 1990s were larger in size and that there were fewer built than the infographic shows.

3.4 The balance between what is accurate and what is understandable

The main problem was not that we did not have the numbers for cabins isolated, but that the presentation took away all our insecurity and explanation. It made it look like we knew something we did not know. News articles were made based on those numbers, and if the nuances are lost in our own presentations, they will surely not be added in a news article.

These errors happened because the balance between what is accurate and what is understandable to the user is a difficult one. More often than not, there will be a pull between the communications department and the statistics departments. To exaggerate, the communications department's main concern is making the statistics discovered and understandable, while the statistics department's main concern is getting the correct information through. Both of these view points are important, and hopefully the collaboration makes the result both covering enough and understandable. In this case, the simplification went to far.

4. Overcomplicated words – Finding the balance between smooth internal communications and inviting external communications

“Can you talk to 240 about doi?” In Statistics Norway, we want it short, and we want it quick. Knowing the difference between “doi” and “di” is important, as the first one is short for “varehandelsindeksen” (Index of wholesale and retail sales) and the latter “direkteinvesteringer” (Foreign direct investment) – two very different statistics.

One of the best examples of short and quick is how we use numbers instead of names for our departments and divisions. The department for Business and environmental statistics is referred to as 400, and the division for housing, property, spatial and agricultural statistics as 426. Yes, this makes sense, but even the very short department name “IT” is rarely in use – instead we say 700.

One of the very first thing a new employee learns, is the number of their own division and department. But aside from their own number, memorizing the numbers of all the other departments and divisions can take some time. In the meantime, keeping up with conversations in meetings can be a challenge. “Which division did you say I should talk to?” “Who did what?”

In Statistics Norway, using numbers for departments and divisions comes in two phases: First you are confused and have to do an extra round of thinking in conversations. Then, when you learn the correct numbers, you become “one of them”, and you too use numbers like you have forgotten there was ever a challenge.

This pattern can be seen in many areas. In the start of the employment, you have two feet on the outside, and most of what is happening needs to be understood and processed. After a while, you have one foot in each camp, and you are an important

resource in the dissemination of difficult information. You know what the users will struggle with, but you also understand the correct meaning. After a few years, many forget this feeling, as they sink deeper and deeper into the terminology and discipline.

Statistics Norway can be a wonderful bubble with many intelligent people, but there is also a “danger” in that. When you surround yourself with internal lingo, you can forget that the user on the outside of the building does not share your perception. Also, sometimes this barrier can be found even within the building. Everybody has their own expertise, and IT and economical statistics can have a professional conversation without understanding each other.

4.1 So, what is the solution?

The clear language project aims to remind everybody within Statistics Norway that if we want to reach the user, and make them understand us, we need to write in a way that they understand.

But explaining complex connections in a short sentence and finding simple definitions for long words is the real challenge. It’s much easier to stick to your own professional vocabulary.

5. Conclusion

Language is hard. Sometimes the best word for the user is not the word that best covers the meaning behind the statistics. Sometimes the perfect word is so long and complex, the user never even looks past the first sentence. In those cases, we in Statistics Norway, and others who work with statistics, need to find a sweet spot – an area where both those with low and those with high competence on the subject feels welcome. Either with that perfect word that covers both the true meaning and makes sense to the user, or with simplification as a starting point and plenty of room for the explanations and nuances further down.

Failing should not keep us from trying and experimenting with new ways to present facts. It should remind us to always look for improvements.

6. Citations

Lovdata (2022, 1) *Lov om offisiell statistikk og Statistisk sentralbyrå (statistikkloven)*. Available at: <https://lovdata.no/dokument/LTI/lov/2019-06-21-32> (Accessed: 20 April 2022).

Lovdata (2022, 2) *Lov om språk (språklova)*. Available at: <https://lovdata.no/dokument/NL/lov/2021-05-21-42> (Accessed: 5 May 2022)

Mathiesen, J. (2022) *Rekordmange byggetillatelse til nye hytter*. Available at: <https://www.ssb.no/bygg-bolig-og-eiendom/bygg-og-anlegg/statistikk/byggeareal/artikler/rekordmange-igangsettingstillatelse-til-hytter-og-fritidsbygg-copy> (Accessed: 28 June 2022).

Statistisk sentralbyrå (2022) *Byggeareal*. Available at: <https://www.ssb.no/statbank/sq/10070025> (Accessed: 24 June 2022)

Store Norske Leksikon (2022) *Hytte*. Available at: <https://snl.no/hytte> (Accessed: 20 April 2022)

7. Illustrations

Illustration 1: <https://en.wikipedia.org/wiki/Half-track>

Illustration 2: <https://www.colourbox.com/image/snow-mobile-in-the-farm-at-winter-rovaniemi-image-29907818>

Illustration 3: Private photo own by Thistel, L.S.

Illustration 4: <https://www.colourbox.com/image/modern-house-in-bright-summer-day-image-14191973>

Illustration 5: <https://www.colourbox.com/image/candanchu-image-11461057>

Illustration 6: <https://www.colourbox.com/image/trail-leading-up-to-a-small-red-cabin-image-29837811>