On passing time

Organising time in language

Marc van Oostendorp

isbn 978-90-71093-97-5

Text edited by Gary Schwartz.
www.kb.nl/fellowlezing
www.nias.knaw.nl/kb-lecture

One interesting property of human beings is that they can feel structure in time. Time passes, one moment after another, and we can feel that not every moment is the same as every other moment. We can also feel that different kinds of moments alternate, and that those alternations follow a pattern. Such patterns are called rhythm.

There are junctures when it feels like time is standing still, because every moment is similar to every other moment. At other times it seems as if everything is moving too rapidly, because every moment seems different from every other one. To the student of rhythm, these kinds of time are not very interesting. The first is rhythmically too simple; the second has no rhythmic structure at all. ‘Real’ rhythms fall somewhere between those extremes.

At the same time, rhythms are always in some respect illusions – structures imposed on reality by our minds. Since two events are never really the same, we exercise a subjective judgement allowing us to perceive as 'identical' things that are no more than ‘similar enough’.

The illusion of rhythm is very strong; indeed, it is essential for many human activities. It helps us to organise time by way of a clock; it helps us to coordinate our activities. Without rhythm, we could not walk, let alone dance, or sing. Fortunately, most people have a sense for rhythm, even if they deny it. To the very few people who really lack such a sense the world is a difficult place. They cannot appreciate music and sometimes may not capture subtleties in human communication.

Rhythm is also an intriguing property of human language. The words and syllables we speak are produced one after the other, but not at random intervals.

Indeed, we know that language that is spoken at random intervals is very difficult to understand. Furthermore, we know that the rhythm of one's native language is deeply entrenched in one's mind. It seems to be among the very first things we learn. Babies only a few days old are able to recognise the rhythm of their mother tongue. They listen to a voice speaking their native language more attentively than when the same person speaks a different language. Even if the signal is manipulated to disguise individual vowels and consonants, the effect is still there.

The fact that children learn this rhythm so early on is an indication of how important it is that they learn their own language, which is after all an instrument for learning everything else in the world. There is evidence that they actually begin to absorb their mother('s) tongue even before they are born. Since individual vowels and consonants might not be easy to distinguish through the walls of the womb, the child will start by learning those things she *can* distinguish: the rhythmic structure of speech, sentence melody and perhaps a few other attributes related to these rhythmic properties, including certain aspects of the way in which sentences are built.

The rhythm of one's native language shows up in other aspects of life as well. The musicologist Aniruddh Patel has shown that there is a subtle but systematic difference between the rhythms of French composers like Debussy and such British contemporaries as Elgar, and that these can be related to rhythmic differences between French and English. Apparently, our concept of rhythm is not general – it is tied to certain specific rhythms of our mother tongues.

-›|‹-

Over the past months, browsing in the immense holdings of the National Library, I found myself wondering whether rhythmic preferences also show up in the use of language. Can we find in poetry and everyday literary usage phenomena like those in music? My attention was drawn to the digital collections in the National Library, which are full of text. Can we find rhythms in these texts? Were poets and writers influenced in the same way by the rhythms of their mother tongue as composers?

The field of linguistics would seem to offer a way to answer that question. Linguists draw a rough distinction between three types of language in terms of the organisation of the time they take to produce a word or a sentence: English is the model of a language that takes stress as the main unit of counting; French stands for the group that favours the syllable; the third group takes a unit smaller than the syllable as a unit of counting.

In languages like French, Italian or Spanish, every syllable is more or less of equal length, i.e. it takes more or less the same amount of time to pronounce. In English, Dutch and like languages, unstressed syllables take much less time than stressed syllables; furthermore, syllables can end in clusters of

consonants that are sometimes quite lengthy. While syllables can therefore be of very unequal length in these languages, the amount of time taken for any two stressed syllables is more or less the same: in the sentence *father walks every day*, the words *father*, *walks*, *every* and *day* each takes more or less the same amount of time, because all have one stressed syllable.

Japanese, finally, is a typical language of the third type. If a syllable is closed (like *nan*), it takes almost twice the time of a syllable that is short (like *na*).

Interestingly, we see that poetic traditions often seem to be based on those principles. For instance, classical French poetry uses types of verse that count syllables, like the so-called *octosyllabe*, in which every verse has exactly eight syllables:

Ah! marquise, quand on y pense,
Ce temps qu'en folie on dépense,
Comme il nous échappe et nous fuit!
Sais-tu bien, ma vieille maîtresse,
Qu'à l'hiver, sans qu'il y paraisse,
J'aurai vingt ans, et toi dix-huit?

(‘Ah! Milady, when you think about it, the time one spends on folly, how it escapes us and flees from us! Do you realise, my aged mistress, that this winter, without looking it, I will reach the age of twenty and you eighteen?’) Alfred de Musset, *A Juana*

Very little English or Dutch poetry uses the same kind of metrical structure. Rather, our poetry is usually built around the so-called verse foot, a combination of a stressed syllable with one or more unstressed syllables. A common foot is the so-called *iamb*, consisting of an unstressed syllable followed by a stressed one. The following poem has six iambs in every line:

Onachterhaalbre Tijd, wiens hete honger graag
Verslokt, verslindt, verteert al wat er sterk mag lijken
En keert en wendt en stort Staten en Koninkrijken;
Voor iedereen te snel, hoe valt gij mij zo traag?

(‘Uncatchable time, whose hot hunger eagerly sucks up, devours, digests everything that may seem strong, which overthrows, reverses and smashes States and Kingdoms; if you are too quick for everybody, why are you so slow for me?’) P.C. Hooft, *Sonnet*

Finally, in Japanese poetry, closed syllables count double. The famous haiku has three lines, consisting of five, seven and five units respectively. In western translations, those units are typically taken to be syllables; but in the Japanese tradition, a closed syllable counts for two:

江戸の雨何石呑んだ時鳥

e-do no a-me (5)

na-n go-ku no-n-da (7)

ho-to-to-gi-su (5)

(‘How many litres of Edo’s rain did you drink, cuckoo?’) Issa, Haiku

The syllables *nan* and *non* in this poem count as two units, so that the second line actually has five syllables.

As suggestive as they are, these examples are of course not statistically significant; there has never been a large-scale typological study, as far as I know, which conclusively links such choices by poets to the rhythmic properties of the languages they speak.

Nonetheless, the individual observations just quoted are based on rather deep analyses of the languages in question.

-›|‹-

What I have sketched is a monumental research project – for somebody else. It would involve a gigantic amount of work. One has to establish for a representative number of languages what their rhythmic structure is, and at the same time how their poetic tradition works.

Each of those tasks would take a lot of time on its own. In order to establish the rhythmic type of a language, one needs to have recordings of the language and means of dividing the sound stream into syllables. Anybody who has ever listened to a foreign language radio station knows that this is next to impossible to do if you do not know the language well.

Furthermore, in order to be able to do it properly, you must also be able to count consonants and vowels. You would have to find people capable of performing this task for each of the hundreds of

languages which – because the differences between languages can be great – are typically required in typological work for a representative sample. As far as I know, iresearch of this kind is only being carried out on a handful of languages, most of them European or related to European languages.

Defining the ‘poetic tradition’ of a language is even more difficult. One cannot be contented by looking at randomly chosen poets, which may create a false impression. This could result from the present-day use of non-metrical, free verse by many Dutch, French and Japanese poets; in past centuries as well, however, we encounter attempts to work in other traditions than the ones I described, even if they ‘failed’ in the sense that the tradition did not catch on.

Although in principle it would therefore be possible to test this hypothesis, the amount of rather specialised work required makes it next to impossible to realise it within a reasonable amount of time. At the same time, there seems no reason not to stick to the hypothesis. It fits the data we know and it makes sense as well. So that is what we will do for the time being.

-›|‹-

What we can do in the meantime is to study in greater detail what we know about languages for which we have gathered enough data of the right quality *and* have acquired enough insight to interpret it.

That is what I set out to do, with regard to the Dutch language, during the six months I spent at the Netherlands Institute for Advanced Study (NIAS) and the National Library of the Netherlands (KB): to see how we can understand more about the rhythms of Dutch and what it can explain about this particular language. Dutch lends itself to this campaign because it has been studied so extensively by previous generations of researchers, and because their work has been preserved and documented in the National Library and other repositories.

A second aim for my tenure as KB Fellow of NIAS was to ponder what it means that we are getting more and more information online, in our case the enormous amounts of Dutch text on the Internet, much more than we will be ever able to read. According to some estimates, the amount of digitised information that had been gathered by mankind from the beginning of time until the year 2003 is no larger than the volume of digitised information that has been

generated during the ten minutes I have been speaking to you. Not all of that information is language, but a lot of it is; and that is something we have to deal with in one way or another.

Discussion about this topic has a tendency to get polarised. There are those who believe that these tools allow us finally, and for the first time in history, to perform objective and truly scientific research on language. They say that we have to move on as quickly as we can. The other party, inversely, would say that these new developments are nothing but a passing fashion craze that can never bring us anything worthwhile.

It will be obvious to most people in the audience that both positions are too extreme, and that the truth lies somewhere in the middle. However, such compromise solutions are rarely sought, and it is indeed not easy to find them. It seems to me apparent nonetheless that the enormous flood of data coming at us, after an initial period in the digital age when we suffered from a scarcity of relevant data, is going to bring about an enormous change in the way we perceive and deal with our object of study. But it also should be clear that any good scholarship is

cumulative, that it starts from knowledge and insight accumulated over long decades.

Even if it is true that we are now gathering more digital data than everybody ever did before, this does not hold for our *understanding* of things. One sometimes gets the impression that the evangelists of digital data believe that such understanding comes automatically once we have accumulated enough material and let the algorithm do the 'analysis', i.e. find correlations between them. Understanding comes at its own pace, and it is difficult to speed it up. The best we can do is to 'stand on the shoulders of giants' and let ourselves be inspired by insights that are already there. In order to make real progress, we have to be familiar with old insights as well as understanding the new tools.

We are thus living in exciting times, but knowing that does not give us an answer to the big question that confronts us.

Early on I realised that I had no idea how to answer that question in general terms. I am not even sure the question is answerable in principle, since it would require looking into the future. The best I can do is to query some of the concrete issues with which I am

wrestling and see how they work out. Take for example my question about rhythm.

Together with a software developer at the National Library, Michel Koppelaar, I produced a small tool for scanning text, that is to say for finding regular patterns of stressed and unstressed syllables. In particular we concentrated on the iambic pattern which I explained before: lines or sentences made up of sequences of syllables that sound like *ta-DAM ta-DAM.*

I wanted to test this tool on three types of data, to see whether it could answer questions of three different kinds.

The first is an historical question. We know that iambic patterns in poetry came to the Netherlands during the Renaissance. They originated with Italian poets, most prominent among them Petrarch, and spread across Europe. It is usually believed that the iamb hit the shores of our Dutch-speaking corner of the continent sometime in the fifteenth or sixteenth century. Before that time, some version of more original 'Germanic' prosody was used, which counted only the number of stressed syllables in a line; unstressed syllables did not matter at all.

That was at least what was thought until a few years ago, when the Dutch phonologist Wim Zonneveld published an article on the early fourteenth century hagiographic text *Sinte Lutgarde*. Read with present- day accentuation, this text sounds entirely iambic. The few exceptions are systematic and understandable. For instance, proper names must be read with the original Germanic stress. (But we know that the pronunciation of proper names tends to be conservative.)

Zonneveld's argumentation is circular in an interesting way. Our main evidence concerning stress in older stages of language lies in metrical poetry. We have no recordings of how words sounded, nor was word stress written down any more frequently than in modern Dutch. We do know, however, that stress in the pronunciation of Dutch must have been different than the Germanic system; in Dutch, stress fell systematically on the first syllable. This changed during medieval times with the wholesale influx of Latin loan words; to this day the Dutch word stress system shows the influence of this linguistic event.

As I hinted, the only evidence we have that stress had taken more or less its modern shape in the time of *Sinte Lutgart* is that the poem is written in iambs. At the same time, we only know that it is

written in iambs because that is what it sounds like if we read it with modern Dutch stress.

In my view, the circle is so strong that it would be difficult to break it. The two arguments really support each other. Yet mysteries remain: why would there be an isolated case of a regular iambic poem in this period, followed by at least a century of absence of such a pattern, and retreat to the older system?

In order to be able to answer this question, we would have to know whether *Lutgart* was really as isolated a case as all that, or whether we just happened by accident on this example of a more common phenomenon. Perhaps other writers from the period were trying to be iambic, even if they did not always succeed.

A useful step that might help to answer that question is to undertake the scansion of a large number of poems. The problem with this kind of work is that it is rather labour-intensive, while requiring quite some knowledge from the researcher, making it more expensive than it seems to be worth. The solution for which we opted was to employ the computer. Automated reading cannot do the entire job – scansions are sometimes subtle and require human understanding – but if the computer can do a fast and

dirty job on a very large number of examples, this would definitely be a great help.

I cannot yet show you all our results, since it turned out to be harder than we thought it would be to deal with the enormous spelling variation in medieval texts. The reason for this is interesting for what it tells us both about the texts and about our method. Our algorithm works as follows. It looks up every individual word in a pronunciation lexicon that tells which syllables in a word are stressed and which vowels are schwa. Unstressed schwas are more unstressed than other unstressed vowels, because a schwa can *never* be stressed. The algorithm reads a line and replaces it with a set of numbers. If a word is unknown, it provides only the number of syllables, with a code indicating that the program does not know the stress value of the sample concerned.

The algorithm then tries to map the line on the ideal template. In our example, this is a fixed number of iambs, but any other model can be substituted as well. This results in a notably permissive interpretation, which probably resembles the practice of most readers.

However, because spelling variation in medieval Dutch is so extreme, combined with the fact that relatively many words are no longer known, our

test yields too many false positives for it to be effective. There are just too many lines that are seen as completely regular even though a human reader might not see them as such. (Our friend the computer will *always* assume that words it does not know have the right pattern.)

Fortunately for us, the second question involved iambs in modern poetry, where we know all the words. In particular, I concentrated on what is probably the most common verse type, the so-called iambic pentameter, which consists of five iambs in a line:

Je raakt de mensen en de dingen kwijt,
tot je het leven langzaam voelt verglijden
en deel wordt van het raadsel van de tijd.

(‘You lose the people and the things, until you feel life slowly leave, and become part of the riddle of time.’) Jean-Pierre Rawie, *Raadsel*

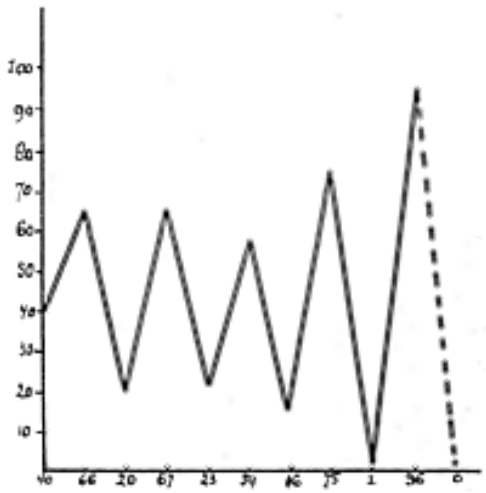
Typically, poets tend to avoid giving every line exactly the same rhythm. This in itself is interesting. Is it considered boring if every line sounds the same?

Why should it? Why do people think that variation is necessary for aesthetic pleasure?

However that may be, in order to achieve this effect, poets allow themselves small and subtle deviations from the pattern. A typical example can be found in the second line of the Rawie terzet above: *tot je* has a clearly unstressed syllable in the second position, where it is supposed to be stressed. One therefore senses a slight stress on the preposition *tot*. This phenomenon is called *iambic reversal*: instead of *teDAM,* the lines starts with *DAMte.*

Iambic reversal is most often found in initial positions. In this typical case it appears at the beginning of the line. There seem to be very few poets in Dutch, English or other modern Germanic languages who do not take this liberty. Some poets do it at other starting positions, for instance the beginning of a sentence that starts halfway in a line. In the poem by Hooft, the mid-line reversal is usually assumed to have some kind of sound-symbolic meaning: *En keert en wendt en stort Staten en Koninkrijken* inverts *Staten,* a reversal that symbolises the falling down of states.

An intriguing question now presents itself: why should lines be asymmetric in this way? This issue preoccupied Dutch linguists and philologists particularly in the 1930s. Research conducted in that period revealed many more asymmetries than had been noticed earlier. For instance, the great phonologist A.W. de Groot, after scanning a large number of lines by hand, came up with graphs like the following:



The horizontal axis represents the ten syllables we might expect in an iambic line. (The dotted line moves to a facultative eleventh syllable for cases of so-called ‘feminine rhyme’, such as *verglijden* in the line by Rawie above.) The vertical axis gives the percentage of such syllables in the corpus that are stressed. The graph clearly shows that the amplitude of the change

enlarges towards the end: syllable #9 is virtually never stressed, whereas syllable #10 almost always is. The latter phenomenon might be due to the fact that these lines rhyme, and rhyming syllables are typically stressed. However, this does not explain why the syllable immediately before it is always unstressed, or why syllable #8 is more likely to be stressed than any other even-numbered syllable except syllable #10.

De Groot's insightful observations about the structure of the line remain valid. In addition to observing the tendency for lines to become more rhythmical towards the end, he put his finger on another, albeit weaker, tendency to have stronger rhythm at the beginning of the line than in the middle. The effect of this is that the sixth syllable, right in the middle, is the least likely to be stressed. Iambic reversal at the absolute beginning of the line is a third tendency. Since it counteracts the tendency for beginnings to be more metrical, the upshot is that the second and fourth syllable have more or less the same amount of stress.

As is always the case in science and scholarship, this explanation asks in turn for further explication. De Groot established the following three principles guiding the structure of the line:

i - Lines are more regular towards the end.

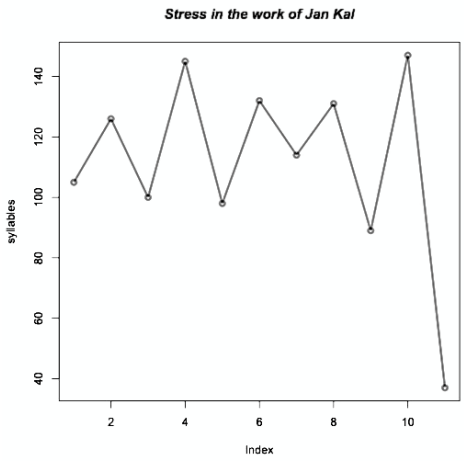
ii - Lines are more regular towards the beginning.

iii - The first foot may undergo iambic reversal.

We now need to clarify each of these principles, and explain why (i) is so much stronger than (ii). Now, as it happens, (ii) and (iii) have a more straightforwardly functional explanation than (i). If the intended rhythm is not established at the beginning of a verse, the reader or listener will not be able to detect it at all. That could explain (ii). It is also important to start with a stronger element to attract attention; that could be the explanation of (iii). But given all this, why should the ends of lines be so regular? What is it that makes this tendency the strongest of all?

That the latter is the case can be objectively demonstrated. In order to test de Groot's work, I repeated it with the algorithm we developed at the National Library. The graphs that emerged are very similar to his. A clear parallel is apparent in the work of Jan Kal, a contemporary author of sonnets. A few years ago Kal published a collection of 1,000 sonnets, and by kind permission of the publisher we were provided with a digital version of that collection,

allowing us to conduct automatised scansion. Within a few minutes we drew the following graph:



As you can see, the result is quite similar to what de Groot found after long and arduous work. The end of the line is very regular, as is the beginning, except that the very first foot shows iambic reversal. This effect seems to be stronger for Kal than for the poets de Groot studied (Perk, Kloos and Boutens, for instance). Furthermore, the irregularity goes beyond the fact that syllables 6 and 8 are stressed less than other even syllables. It also extends to syllable 7, which is stressed *more* than the average for odd-numbered syllables.

So the effect seems to be real. I found it for many other poets as well, and did not encounter any author who shows a clearly different pattern of

exceptions. In sum, then, we can conclude that behind de Groot's observations lies a reality of some kind.

-›|‹-

That brings us quite a way along in our investigations, but it does not explain why these restrictions exist in the first place. One circumstance that might be relevant is that music seems subject to a similar restriction. For instance, a cadence in a concerto or an improvisation in a jazz number will typically start more or less in line with the written music and then become looser, to revert to the beat towards the end.

Spoken language sentences too display certain left-right asymmetries. They tend to slow down towards the end, and in most languages the tone drops lower as well. An exception to this rule is that questions are usually asked in a raised tone throughout the sentence, up to the end. These asymmetries seem universal, as can be shown by a simple experiment. Play a text sentence backward in any human language and these characteristics would alert an attentive listener to the reversal even if he or she does not know the language in question.

These findings lead us to conclude that our way of dealing with time is asymmetrical. The beginning is different from the end. This finding too demands explanation, and various reasons have been advanced. That we lower our voices towards the end of a sentence, for instance, is sometimes explained by the fact that we speak by breathing out, leaving us with less energy towards the end of a sentence than at the beginning. This explanation is based on the behaviour of our bodies. Indeed, we need our bodies to give verbal expression to sentences, whether or not in verse, and our bodies move through time in an asymmetric way, from the past towards the future.

Yet I doubt whether the asymmetry in question can be linked directly to the body. Metrical structure is abstract. It belongs to the mind, and has no obvious physical correlative.

As I see it, language is abstract at the deepest levels of its structure. There is no ‘left-hand side’ and ‘right-hand side’ in the *meaning* of *John smokes,* there is only the relation between the subject and the verb. The fact that one precedes the other in English has no bearing on the meaning of the sentence. In other languages the order might be reversed, but the meaning stays the same. The fact that *John* precedes

smokes when we speak is only the result of the fact that our mouths cannot *say* two words at the same time. It is mainly in this specific way that the body can be considered the decisive agent.

-›|‹-

What, then, are we dealing with in essence? What in the world makes sentences and music asymmetric? One hypothesis reduces the question to a matter of coordination. We do not pronounce sentences in isolation or even in a monologue. We very often speak them in dialogue, and together. Dialogues are great coordination games. Turns are taken very smoothly: in most cases, one person starts to speak more or less exactly at the moment when the other ends his contribution. Even interruptions seem timed. To carry them out completely at random would create the impression that the interlocutor was not listening at all.

We thus have a feeling for the right moment to start to speak. Perhaps we pick up a signal from the speaker, who by slowing down and changing his tone of voice signals that he is approaching the end of his

turn. The listener can then start preparing the next thing to say: formulating the thought, finding the right words, putting them in the right order, sending instructions to his tongue and lips, and performing all these other complicated, time-consuming tasks.

The familiar properties of the end of a sentence may be explainable in this way, but what about the rhythmic properties of poetry - and music, for that matter? Although we may be excused for thinking that rhythmic tendencies in music and language have the same origin, no study seems ever to have been devoted to the subject. The next step that I will be taking with Michel Koppelaar is to see whether we encounter the same effects in theatre plays as in poetry. We are currently in the process of testing this.

This question leads us back to our original programme, that is whether the preferences identified by de Groot and us in poetry can also be found in prose texts. The National Library currently has a large number of prose texts in digital form, and offers access to many more. We are analysing these writings in an attempt to see whether or not they too, like poems, follow rhythmic patterns. In particular, the relative popularity of iambs in poetry is sometimes explained by positing that ‘ordinary Dutch sentences

are also iambic', since they start with an unstressed word like *de*, followed by a word with stress on the first syllable, like *moeder*. The claim that non-poetic Dutch, like Dutch poetry, also tends towards the iambic has never been empirically verified, however. With our new tool, it will be easy to prove or disprove it.

Once calibrated, our tool can then be used to locate other places where the iambic preference might show its face. For instance, in Dutch the word order of verb clusters at the end of the phrase is relatively free. There is not much of a difference between the following two sentences, which both have the same meaning:

- *Ik denk dat jij een boek gelezen hebt.*
- *Ik denk dat jij een boek hebt gelezen.*

(I think that you have read a book.)

The choice between the two sentences is influenced by all kinds of factors: in some regions one might be used more frequently than the other, for instance. Some people might think that the first should be avoided in Dutch for the sole reason that it is the (only) correct

one in High German. Some scholars think that rhythm plays a role. The first phrasing, they point out, ends in an iambic pattern, and the second one does not. In the plural the same sentence has the opposite properties:

- *Ik denk dat jij boeken gelezen hebt.*
- *Ik denk dat jij boeken hebt gelezen.*

(I think that you have read books.)

Recently, the Flemish linguist Georges De Schutter has argued that a tendency to become iambic might indeed play a role in the way sentences are built. The idea is that since *gelezen* (the past participle of 'read') starts with an unstressed syllable, the singular *boek*, which has only one stressed syllable, can precede it quite easily. However, the plural form *boeken* also ends in a stressed syllable, so putting it immediately before *gelezen* would create a 'lapse' of two unstressed syllables in a row. If De Schutter is right, people would avoid such lapses as well as the clash that would arise if the singular *boek* were to be followed immediately by the equally stressed monosyllable *hebt.*

Even if De Schutter is right, however, the rule can only identify a statistical tendency; as we noted, multiple factors must be involved. As far as I am able to determine, those other factors do not take into account the difference between singular and plural. We intend to use our tool to research this question as well. (The present tool is not fitted out for this problem, because it does not recognise the syntactic structure of sentences; we need another instrument that *can* do this, which would then be implemented to see whether alternatives are available.)

-›|‹-

This completes my short description of the kinds of research we are doing right now. Clearly, we are still in the middle of things; exploring the world is an enterprise that never ends. What we want to know at this interim phase is: what does our campaign teach us about the digital humanities?

The projects just sketched clearly fall in this category. Depending as they do on digital data, preferably great masses of them, they can safely be called big data projects in the humanities. What really

interests us is how they differ from other big data projects, which they do in a number of ways.

In the first place, they build on existing theories, some of which were developed many decades ago. They even use methodologies that are not essentially different from those of earlier periods. In this sense, we challenge the idea that science progresses through revolutionary new insights into the object of study. If there is anything revolutionary about our work, it lies in our new ability to apply older methodologies more effectively. Our predecessors spent long years pondering their aims, but they were not able to pursue them successfully in a reasonable amount of time. Like us, they too came up against technical limitations, albeit on a smaller scale.

The introduction of digital methodologies is clearly a boon to the study of the relation between rhythm types in languages of the world and their respective poetic traditions.

I believe that research of this kind makes more sense than the kind of ‘data mining’ that is becoming more and more popular these days. The philosophy behind data mining is that specialist knowledge concerning the subject of research is undesirable on the part of the researcher. It introduces bias and is 'not objective'.

An example of data mining is the kind of research that seeks to define the features that distinguish a literary work from non-literary writing. It is quite widely accepted that literary quality is a function of two forces: a sociocultural one (there has to be an elite deciding what counts as literature) and a formal one (something in the text itself which makes it valuable). The ratio of these two factors and the way they interact are less clear, but that has not prevented a host of researchers, convinced of the primacy of one side or the other, from devoting a lot of work to the issue in the course of several centuries.

Quite bizarrely, some current research projects have started to explore the question all over again by collecting large amounts of the most extreme examples of both kinds and trying to correlate them. One way of going about this is to ask a large group of people about their opinion concerning an equally large number of books, and then to correlate the out-come with aspects of the highest scoring works that are easily quantifiable, such as the percentage of adjectives or the average length of paragraphs. A methodological weakness of this technique is that if one takes a sufficient number of parameters into consideration, one is bound to find correlations here or there.

In my view, the discovery of such correlations does not add to our insight, if only because they probably cannot be embedded into what we already know about how literature works. The correlations themselves are just random new facts. We can accumulate them, but why should we? There is no shortage of facts, and no need to produce random new ones. What is lacking is insight and understanding. While we surely can use computers not only for number crunching but also to improve understanding, in order to do so they have to be used in a more sophisticated way and applied to very specific questions in the material.

-›|‹-

In the last few months, I have been fortunate enough to be allowed to think about how we organise time and the best questions to ask about the organisation of time in the present day and age. To add to my good fortune, I was able to do so in two exceptional locations that are symbols of timelessness while finding themselves in a state of transition.

The **Netherlands Institute for Advanced Study in the Humanities and Social Sciences** (NIAS) has for several decades provided peace, quiet and inspiration for scholars to produce great works. Unfortunately, it is experiencing some turbulence now, but I am confident that quieter times will soon return. The **National Library** (KB) is also going through various transformations. One important development is that the KB is tackling the transition to a digital era, in what seems to me to be a wise way.

We are all facing an uncertain future, a truth that may
be the single most crucial asymmetry in time
that we humans experience. All the same,
we manage to discover patterns
in the structure of time,
and take pleasure in
it.

www.kb.nl/fellowlezing

www.nias.knaw.nl/kb-lecture