



LOQUENS 1(1)

January 2014, e002

eISSN 2386-2637

doi: <http://dx.doi.org/10.3989/loquens.2014.002>

# Introducing a scansion machine for Dutch poetry and prose

Marc van Oostendorp

Leiden University Centre for Linguistics

e-mail: [marc@vanoostendorp.nl](mailto:marc@vanoostendorp.nl)

**Citation / Cómo citar este artículo:** van Oostendorp, M. (2014). Introducing a scansion machine for Dutch poetry and prose. *Loquens*, 1(1), e002. doi: <http://dx.doi.org/10.3989/loquens.2014.002>

**ABSTRACT:** This paper looks into the possibility that the rhythm present in poetry is the same that characterizes natural languages. This hypothesis, although attractive, faces serious problems such as the tremendous amount of data necessary to reach any valid conclusion and the fact that certain poetic traditions may deviate from the rhythm expected in a certain language. In order to tentatively test whether there is a relationship between the rhythmic properties of a certain language and its poetry we selected Dutch, a stress-timed language, and we took modern Dutch metric poetry as the object of study. Then, we created a tool that finds regular patterns of stressed and unstressed syllables in iambic pentameters and we compared our results to those of de Groot (1936). Finally, we discussed what factors could influence the noticeable findings for the beginning and the end of the studied verses and whether these findings could be the same for Dutch prose texts.

**KEYWORDS:** stress-timed rhythm; Dutch; poetry; scansion; iambic pentameter

**RESUMEN:** *Una máquina de escansión para poesía y prosa en neerlandés.* - En este artículo examinamos la posibilidad de que el ritmo presente en la poesía sea el mismo que caracteriza a las lenguas naturales. Esta hipótesis, aunque atractiva, se enfrenta a importantes problemas, tales como la necesidad de disponer de una ingente cantidad de datos para así obtener conclusiones válidas, o el hecho de que ciertas tradiciones poéticas pueden desviarse del ritmo esperado en una determinada lengua. Con el fin de examinar provisionalmente si existe una relación entre las propiedades rítmicas de una lengua concreta y su poesía, elegimos el neerlandés, una lengua acentualmente acompañada, y como objeto de estudio tomamos su poesía métrica moderna. Seguidamente creamos una herramienta que encuentra patrones regulares de sílabas tónicas y átonas en pentámetros yámbicos y comparamos nuestros resultados con los conseguidos por de Groot (1936). Por último, analizamos qué factores pueden influir sobre los hallazgos percibidos al comienzo y al final de los versos estudiados, y si estos resultados podrían ser los mismos para textos prosísticos del neerlandés.

**PALABRAS CLAVE:** ritmo acentual; poesía; neerlandés; escansión; pentámetro yámbico

## 1. INTRODUCTION

Rhythm is an intriguing property of human language as well as of a variety of other activities, ranging from music to walking. The words and syllables we speak are produced one after the other, but not at random intervals. Rhythm, therefore, is an important object of study for the speech sciences.

Furthermore, we know that the rhythm of one's native language is something that is very deeply entrenched in one's mindset. It seems to be among the very first

things we learn, for instance. Babies just a few days old are already able to recognize the rhythm of their mother tongue. If they listen to the same speaker speaking their native language, they listen more attentively than when the same person speaks a different language. If we manipulate the signal in such a way that individual vowels and consonants cannot be recognized, the effect is still there (Gómez et al., 2014; Langus & Nespor, 2013).

The reason why children learn this rhythm so early, is presumably that it is very important to learn one's language as soon as possible, as this will then soon start

functioning as an instrument for learning everything else in the world. And because individual phonological segments might not be so easy to distinguish through the walls of the womb, the child will start by those things which he or she can distinguish: the rhythmic structure, as well as sentence melody and maybe a few other properties. Linguists have also shown that some other properties might be correlated to these properties, including certain aspects of the way of syntax, in particular word order (Shukla & Nespors, 2010), so that we learn also those in the womb.

It has furthermore been shown that the rhythm of one's native language may also show up in other aspects of cognitive behaviour. Patel and Daniele (2003), for instance, have shown that there is a systematic difference between French composers like Debussy and their British contemporaries like Elgar in subtleties in their rhythms, and that these can be related to rhythmic differences between French and English. Apparently, when we think of rhythm, we think of certain specific rhythms that are related to our mother tongues.

## 2. RHYTHMIC PROPERTIES OF POETIC TRADITION

One question is whether rhythmic preferences also show up in the artistic use of language, both in poetry and in styles of writing that come closer to everyday usage. Could we also find rhythms in these texts? Would poets and writers be influenced in the same way by the rhythms of their mother tongue as composers?

There seem to be some indications in this direction. Linguists roughly distinguish between three types of language (Nespors, Shukla & Mehler, 2011): those, like French, which take the syllable as the unit of counting; those, like English, which take stress as the main unit of counting; and those which take a unit smaller than the syllable as a unit of counting, that is, the *mora*.

In languages like French (or other Romance languages), every syllable is more or less of equal length, which means that each syllable takes more or less the same amount of time. On the other hand, in a language like English or Dutch, unstressed syllables take much

less time than stressed ones, and furthermore, syllables can end in sometimes lengthy clusters of consonants. This means that syllables are of very unequal length, but the amount of time between two stressed syllables is more or less the same in these languages: in the sentence *father walks every day*, the words *father*, *walks*, *every* and *day* each take more or less the same amount of time, because they have one stressed syllable.

Japanese, finally, is a famous language of the third type. In this language, if a syllable is closed (like *nan*), it takes almost twice the time of a syllable that is short (like *na*). The unit of counting in this case is usually called the *mora*. The three language types are typically referred to as syllable-, stress-, and mora-timed languages, respectively.

Interestingly, we can see that poetic traditions often seem based on those principles (Fabb & Halle, 2008, 2012). 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. The following are a few lines from Alfred de Musset's "A Juana" (1810-1857):

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?<sup>1</sup>

There is little successful English or Dutch poetry that uses the same kind of metrical structure. Rather, poetry in West-Germanic languages is usually structured around the so-called verse foot: a combination of a stressed syllable with one or more unstressed syllables. A very common foot is the so-called *iamb*, consisting of an unstressed syllable followed by a stressed one. The following poem, "Sonnet", by the Dutch Renaissance poet P.C. Hooft (1581-1647) 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?<sup>2</sup>

<sup>1</sup> 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?

<sup>2</sup> 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?

Finally, in Japanese poetry, closed syllables count double. The famous haiku have 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, such as in the following famous example by Kobayashi Issa (1763-1828):

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)<sup>3</sup>

The syllables *nan* and *non* in this poem count as two moras, so although the second line actually has five syllables, it nonetheless has seven moras, which is what counts.

These examples are suggestive: a syllable-timed language like French uses syllables, a stress-timed language like Dutch builds its meter based on stress instead, and mora-timed Japanese also counts moras in its poetry. But obviously they are far from 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. We just have these individual observations, which are however based on rather deep analyses of the languages in question.

This is therefore a research project which somebody could take up; but notice that 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 also how their poetic tradition works. One would, in other words, need to connect typological databases on rhythmic type, and on poetic traditions. Neither of those two things seems very straightforward to do, and neither has been done.

As to the linguistic part, work has been done so far on a handful of languages, and as far as I am aware, this work has not resulted in a typological database. The *World Atlas of Linguistic Structures* (Dryer & Haspelmath, 2013) does not have a chapter on rhythmic typology. (Or to be more precise, there is a chapter on “rhythm types”, but it talks about something different, namely the typology of secondary stress in word accentuation.)

Setting up a database of poetic traditions would be even more difficult. One cannot be contented by only looking at some randomly chosen poets. Presently, many Dutch, French and Japanese poets write non-metrical, free verse; in past centuries, they have tried to work in other traditions than the ones I described, which sometimes “failed” in the sense that the tradition did not catch on (there have been attempts of writing syllabic verse in the Netherlands, but this never really caught on). It is unclear how each of these should be interpreted, and if we encounter a new culture, whether we would not

sometimes find one which happens to have the “wrong” poetic tradition at the time.

So, although in principle it would be possible to test this hypothesis, the amount of rather specialized work which would need to be done makes it nearly unfeasible, at least within a reasonable amount of time. However, there seems to be no reason not to stick to the hypothesis, since it apparently fits at least the data we know. Consequently, that is what we will do for the time being.

### 3. USING ELECTRONIC DATA TO STUDY THE DUTCH TRADITION

What we can do, in the meantime, is to study more in detail what we know about individual languages for which we have gathered enough data of the right quality and developed enough insight.

The remainder of this paper will concern itself with understanding more about the rhythms of Dutch, and what they can explain about what we see in this particular language, which has been well studied thanks to many previous generations of researchers, and which has been well documented. Together with a software developer at the Dutch 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<sup>4</sup>. The current essay should be seen as a brief presentation of that tool and its (potential) applications.

The tool can be used in a variety of ways, and in order to discover all kinds of rhythmic patterns in all kinds of written Dutch texts. The algorithm works as follows. It looks up every individual word in a pronunciation lexicon, which marks also which syllables in a word are stressed, but for instance also which vowels are schwas. Unstressed schwas are more unstressed than other unstressed vowels, because a schwa can never be stressed in Dutch. So what the algorithm does is that it takes a line and replaces it with a set of numbers. If a word is unknown, it just marks the number of syllables and gives each of these syllables a special code that it does not know the stress value of the language. The algorithm then tries to map the line as well as it can on the ideal template (i.e., in our example, a fixed number of iambs, but it can be anything). This means it takes the most permissive interpretation: that is what we probably do as readers as well.

The default for the latter input type is the iambic pentameter because that is what I have been most interested in: lines or sentences which are made up of sequences of syllables in which five stressed syllables occur, each preceded by exactly one unstressed syllable. I wanted to test this tool on two different types of data, to see whether it would be possible to answer two different types of questions.

<sup>3</sup> How many litres of Edo’s rain did you drink, cuckoo?

<sup>4</sup> The tool can be found at <http://lab.kbresearch.nl/analyze/Scansion>.

The scansion software is presently built for Dutch, but it does not take much to convert it to a machine that can work for any (stress-timed) language. The only thing it needs is a list of words in standard orthography and phonetic transcription, and a set of rules about which syllable types count as “long” and/or “heavy” in the language in question. Once such a dictionary is provided, the scansion machine will happily churn out scansions also in the new language. The scansion algorithm is open source; the (Python) code can be found at the URL provided (see footnote 4 above).

In the following sections I give two examples of possible applications; their details will be published in more detail elsewhere. My current goal is to show how software of this type can give us interesting phonological insights when we look at text-based corpora.

### 3.1. Asymmetries in modern poetry

The first question involves modern poetry that is written in iambs, which is the most common foot type in Germanic poetry since the 17th century (at least for “artistic” poetry; a lot of “folk” poetry is written in trochees instead, in itself an intriguing observation; see Kazartsaev, 2010; Zonneveld, 1992, 1993). Here we obviously did not encounter this problem. 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, such as in Jean-Pierre Rawie’s “Raadsel”:

Je raakt de mensen en de dingen kwijt,  
tot je het leven langzaam voelt verglijden  
en deel wordt van het raadsel van de tijd.<sup>5</sup>

Typically, poets tend to avoid every line having exactly the same rhythm, something which is in itself interesting. It is considered boring if every line sounds the same, but why is that the case? Why is variation necessary for aesthetic pleasure?

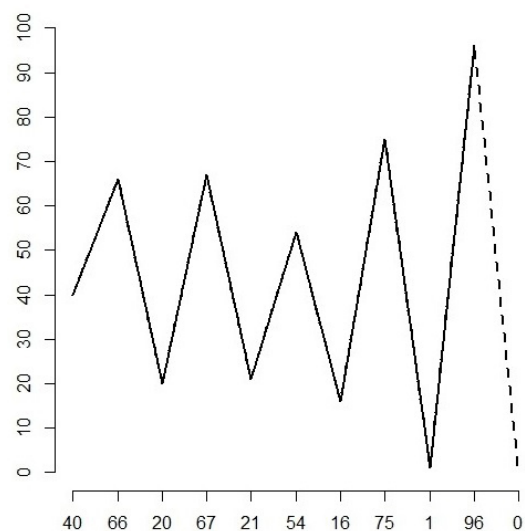
In any case, in order to achieve this effect, poets allow themselves small and subtle deviations from the pattern. A typical example can be found in the second line of the Rawie *terzet* above: *tot je* has a clearly unstressed syllable in the second position, where it is supposed to be stressed. One therefore feels a little stress on the preposition *tot*. This phenomenon is called *trochaic inversion* (Hayes, 1988): instead of with an iamb, the line starts with a trochee.

Trochaic inversion is a phenomenon that is linked to initial positions in verse. In this case it appears at the beginning of the line, and that is a very common phenomenon. There seem to be very few poets in Dutch (or English, or other modern Germanic languages) who do

not take this liberty. Some poets might also do it at other beginnings, for instance the beginning of the sentence, even if this sentence starts somewhere halfway in the line. In the poem by Hooft (see section 2), it is done in the middle in the line, and it is usually assumed that this has some kind of sound-symbolic meaning: *En keert en wendt en stort Staten en Koninkrijken* has inversion in *Staten*, and that symbolizes the falling down of states.

One intriguing question now is: how come that lines are asymmetric in this way? This was a research question for Dutch linguists and philologists in the 1930s. They showed that there were many more asymmetries in the line. For instance, the great phonologist A.W. de Groot (1936) scanned a large number of lines by hand, and in this way managed to draw graphs for the poetry of Dutch poets at the end of the 19th century, such as this one for the poem “Aan Moeders Gral”, by J.L. de Haan:

Figure 1: Stress in “Aan Moeders Gral” according to de Groot. Adapted from de Groot (1936, p. 201).



The horizontal axis presents the 10 syllables we might expect in an iambic line (the dotted line moves to an optional 11th syllable for cases of so-called *feminine rhyme*, such as *vergljden* in the line by Rawie above). The vertical axis gives the percentage of such syllables in the corpus that are stressed. What one can clearly see in this graph is that the amplitude of the change becomes bigger towards the end: the ninth syllable is virtually never stressed, whereas the 10th syllable almost always is. The latter might be due to the fact that these lines also rhyme, and the rhyming syllable is typically stressed, but this does not yet explain why the syllable immediately before it has to be unstressed, or why the eighth syllable is more likely to be stressed

<sup>5</sup> You lose the people and the things,  
until you feel life slowly leave  
and become part of the riddle of time.

than any other even-numbered syllable, except the 10th syllable.

De Groot's observations about the structure of the line are still quite insightful. He observes that there is another, but weaker, tendency to have a strict rhythm at the beginning of the line. The effect of this is that the sixth syllable, right in the middle, is the one which is the least likely to be actually stressed. Iambic reversal at the absolute beginning of the line is a third tendency. Because it counteracts the tendency for beginnings to be more metrical, the end result is that the second and fourth syllables have more or less the same amount of stress.

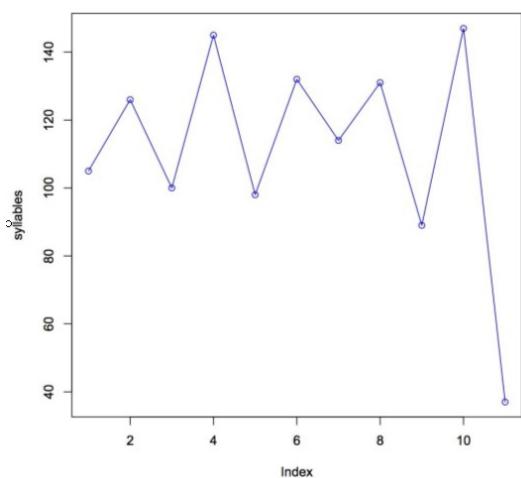
As is always the case in science and scholarship, this explanation in turn of course asks for a further explanation. De Groot established the following three principles guiding the structure of the line:

- (i) lines are more regular towards the end
- (ii) lines are more regular towards the beginning
- (iii) the first foot may undergo trochaic inversion

We would now need to explain each of these principles, and also why (i) is so much stronger than (ii). Now, as it happens, (ii) and (iii) seem to have a more straightforward functional explanation than (i). In the beginning of the verse, the rhythm needs to be established, or otherwise 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 would there be a tendency to also make the ends of lines more regular? And why would this tendency be much stronger?

That the latter is the case can be objectively determined. I have repeated de Groot's work with the scansion algorithm. The graphs we can draw are actually very similar. Here is one which we got for the work of Jan Kal, a contemporary author of sonnets. Kal published a collection of 1,000 sonnets a few years ago, and with kind permission of the publisher we used a digital version of that collection to produce an automatic scansion:

**Figure 2:** Stress in the work of Jan Kal.



It is quite similar to what de Groot found, in his case after long and arduous work and still with much fewer data. The end of the line is very regular, as is the beginning, except that the very first foot shows iambic reversal. This reversal is apparently stronger for Kal than for the poets which de Groot studied. Furthermore, one can also see that the irregularity does not just mean that the sixth and eighth syllables are stressed less than other even syllables, but also that the seventh syllable is stressed more than odd syllables are on the average. (The fact that the 11th position is extremely low is because this syllable is often completely absent, and it was counted as being unstressed.) So the effect seems to be real and we can apparently find it for other poets working in other traditions as well (Fabb, in press). Furthermore, I did not find any Dutch author who shows a clearly different pattern of exceptions—and we can thus conclude that there must be something real about de Groot's observations.

But that does not explain exactly how we got there. One observation we can make (V. deCastro-Arrazola, personal communication, January 14, 2014) is that music seems subject to a similar restriction. For instance, a cadence in a concerto or an improvisation in a jazz jam session will typically start more or less in line with the written music and then become looser; but towards the end it will start being on the beat again.

Also, spoken language sentences display several left-right asymmetries. For instance, sentences tend to become slower towards the end, and in most languages the tone becomes lower, except that questions are usually expressed by raising the tone, also towards the end of the sentence. These asymmetries seem universal. If one played backwards a text's sentence in any human language, one could probably observe this due to such characteristics, even if one does not know the language in question.

We have thus reached another aspect of our way of dealing with time. It is asymmetric. The beginning is different from the end. The question is why this is the case. There may be various reasons. The fact that people's voice gets lower towards the end of a sentence, for instance, is sometimes explained by the fact that we speak by breathing out, and towards the end of the sentence there is less energy. This explanation is thus about our bodies, which move through time in an asymmetric way, from the past towards the future, and sentences are spoken by bodies, as are verses.

Yet the asymmetry in question cannot be directly linked to the body, it seems to me. Metrical structure is abstract. It is something of the mind, and has no clear direct physical correlate.

At the same time, it seems to be the case that language is abstract at the deepest levels of its structure, let us say the way we think: 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 is not about the meaning of the sentence. In other lan-

guages 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. Again, it is the body which decides.

So what is going on in our case? What in the world makes sentences and music asymmetric? One hypothesis is that it is a matter of coordination. We do not just say sentences in isolation or even in a monologue, we very often speak them in dialogue. 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. And even interruptions seem timed: making them at completely random points creates the impression of not listening at all.

We thus have a feeling for when the right moment has come to start to speak. This may be something we pick up from the signal: by slowing down and changing the tone of voice, the speaker may be signaling that he is now 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 to pronounce all those words, and all these other complicated things which need some time.

The familiar properties of the end of the sentence can all be understood in this way, but what about the rhythmic properties of poetry (and music)? One can at least imagine that they have the same origin, although as far as I have been able to find out, these things have never been studied (see Fabb, in press for a different possible explanation). The next step in this line of research is going to be whether we see the effect for instance more towards the ends of clauses than in the beginning in theatre plays. We are currently in the process of testing this extended hypothesis.

### 3.2. Metrical patterns in prose

The previous question also leads us to the second one which was in our original programme, namely whether we see the same preferences also in prose texts. Given that there seems to be an intimate connection between natural language and poetic rhythm, we can wonder whether people follow rhythmic patterns also in prose. In particular, the relative popularity of iambs in poetry is sometimes explained by pointing out the fact that ordinary Dutch sentences are also iambic, since they start with an unstressed word like *de* (the), followed by a word with stress on the first syllable, like *moeder* (mother). This claim has never been empirically verified, however. With our new tool, it will be easy to do so.

This new tool can then be used in other potential ways in which 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, both of which translate into English as “I think that you have read a book”:

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

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 dis-preferred in Dutch for the sole reason that it is the (only) correct one in High German. Some scholars have thought that maybe rhythm also plays a role. The first sentence ends in an iambic pattern, but the second one does not. This would be the opposite for the following pair whose meaning is “I think that you have read books”:

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

Recently, the Flemish linguist Georges De Schutter (2012) has argued that the tendency to become iambic might play a role in the way sentences are built (and not more general rhythmic principles as earlier researchers had assumed). This would only be a statistical tendency, because, as noted, other factors play a role as well.

Together with Marten van der Meulen I developed an extension of the scansion script which takes a prose text and detects verbal clusters like the ones just given. It then reverses the word order of those clusters, and checks which of the two orders gives a better alternation of stressed and unstressed syllables.

The results are quite difficult to interpret (Meulen, 2014). On the one hand, it turns out that for the six Dutch novels which were studied, the rhythm would become significantly less regular if we switched around all auxiliary-lexical verb (Aux-V or V-Aux) combinations. However, at the same time, it turned out that there was an asymmetry between the two orders: if we turned around all original V-Aux orders, the rhythmicity would actually improve. The fact that the word order that was chosen by the authors was highly rhythmic therefore is potentially merely the result of the fact that V-Aux is more rhythmic and it is also chosen more often. The latter might have an independent reason, although it is of course tempting to surmise that the rhythmicity factor plays a role in the preference for this structure.

## 4. CONCLUSION

This completes my short description of the kinds of research we are doing at this moment. Clearly, we are still right in the middle of it: exploring the world is an enterprise that never ends. The question now is what this exploration teaches us about Digital Humanities.



The projects just sketched clearly fall into this category. They crucially use digital data, and actually preferably great amounts: they are big data projects. However, they are also clearly different from other big data projects in a number of ways.

In the first place, the projects sketched in this paper clearly build on existing theories, some of which were developed many decades ago. As a matter of fact, they even use methodologies which are not essentially different from those of earlier periods. In this sense, they thus defy an idea we might have of scientific progress, in which we constantly experience revolutionary new insights in our subject matter. But on the other hand, it also seems that we can do those things only now, after many years of careful thinking by our predecessors, who nonetheless were not able to accomplish such tasks in a reasonable amount of time, in the same way as we are currently unable to reasonably check our hypothesis about the relation between rhythm types in the languages of the world and their respective poetic traditions.

I believe that this kind of research makes more sense than the kind of “data mining” that is becoming more and more popular these days, and that is based entirely on the assumption that it is better to do research without too much established knowledge; such knowledge is seen as a bias and thus “not objective”.

An example of data mining is the kind of literary research which tries to find out what makes a work a literary work. It is widely accepted that there are two dimensions to literary quality: a sociocultural one (there has to be an elite deciding what counts as literature), and a formal one (there is something in the text itself which makes it valuable). The precise mixture of these two factors and the way they interact are less clear, but a lot of work has been done on both sides by researchers in the course of several centuries.

Quite bizarrely, there are now research projects which start all over again in exploring these issues, by collecting large amounts of data on the both extreme sides of the issue and trying to correlate them. One can for instance ask a large group of people about their literary judgement of an equally large number of books, and then correlate this with aspects of those works that are easily quantifiable, such as the percentage of adjectives or the average length of paragraphs. As long as we take into consideration enough of these parameters, one is definitely able to find some correlations there.

But in my view, finding such correlations does not equal insight, simply because they most probably cannot be embedded into what we already know about how literature works. The correlations themselves become just some random new facts, which we can accumulate. There is no shortage of facts, and no need to produce random new ones. There is always a shortage of insight and understanding. We can definitely use computers to get more correlations, but then they have to be used as more sophisticated tools to ask very specific questions to the material.

## 5. NOTE

The present article is an adapted and somewhat extended version of talk given by the author at the National Library of the Netherlands, The Hague, on January 16, 2014.

## REFERENCES

- De Schutter, G. (2012). De werkwoordelijke eindgroep en nog steeds geen einde? [The verbal endgroup and still no end?] *Verslagen & Mededelingen van de KANTL [Reports & Messages from the KANTL]*, 122(1), 1–38.
- Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The World Atlas of Language Structures Online*. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info>
- Fabb, N. (in press). The verse-line as a whole unit in working memory, ease of processing, and the aesthetic effects of form. *Royal Institute of Philosophy Supplement*, 75.
- Fabb, N., & Halle, M. (2008). *Meter in poetry: A new theory*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511755040>
- Fabb, N., & Halle, M. (2012). Counting in verbal art. In I. Jaén & J. J. Simon (Eds.), *Cognitive literary studies: Current themes and new directions* (pp. 163–182). Austin: University of Texas Press.
- Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespors, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 5837–5841. <http://dx.doi.org/10.1073/pnas.1318261111>
- de Groot, A. W. (1936). De structuur van het vers [Verse structure]. *De Nieuwe Taalgids [The New Language Guide]*, 30, 197–212. Retrieved from [http://www.dbnl.org/tekst/\\_taa008193601\\_01/\\_taa008193601\\_01\\_0029.php](http://www.dbnl.org/tekst/_taa008193601_01/_taa008193601_01_0029.php)
- Hayes, B. (1988). Metrics and phonological theory. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey: Vol. 2. Linguistic theory: Extensions and implications* (pp. 220–249). Cambridge, UK: Cambridge University Press.
- Kazartsev, E. (2010). Nederlands en Duits versritme in de vroegmoderne tijd [Dutch and German verse rhythm in early modern history]. *Neerlandistiek.nl*, 10(01). Retrieved from <http://dspace.library.uu.nl/bitstream/handle/1874/190508/article.pdf?sequence=2>
- Langus, A., & Nespors, M. (2013). Language development in infants: What do humans hear in the first months of life? *Hearing, Balance and Communication*, 11(3), 121–129. <http://dx.doi.org/10.3109/21695717.2013.817133>
- Meulen, M. S. van der (2014). Het ritme van rood en groen. Unpublished research report, Leiden University, the Netherlands.
- Nespors, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1147–1159). Malden, MA: Wiley-Blackwell.
- Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87(1), 33–45. [http://dx.doi.org/10.1016/S0010-0277\(02\)00187-7](http://dx.doi.org/10.1016/S0010-0277(02)00187-7)
- Shukla, M., & Nespors, M. (2010). Rhythmic patterns cue word order. In N. Erteschik-Shir & L. Rochman (Eds.), *The Sound patterns of syntax* (pp. 174–188). Oxford, UK: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199556861.003.0009>
- Zonneveld, W. (1992). *Van Afflighem en Chaucer: Het leven van Sinte Lutgart als jambisch gedicht [Van Affliche and Chaucer: Iambic poem of the life of Saint Lutgart]*. Amsterdam, the Netherlands: Stichting Neerlandistiek VU.
- Zonneveld, W. (1993). 700 jaar Nederlandse klemtoon (en weinig veranderd) [700 years of Dutch stress (and little has changed)]. *Spektator*, 22, 198–222.