Chapter 4
The University Was Still Taking Account of *universitas scientiarum*: Wilhelm Ott and Julianne Nyhan

Abstract This oral history interview between Wilhelm Ott and Julianne Nyhan was carried out on 14 July 2015, shortly after 10am, in the offices of *pagina* in Tübingen, Germany. Ott was provided with the core questions in advance of the interview. He recalls that his earliest contact with computing was in 1966 when he took an introductory programming course in the *Deutsches Rechenzentrum* (German Computing Center) in Darmstadt. Having become slightly bored with the exercises that attendees of the course were asked to complete he began working on programmes to aid his metrical analysis of Latin hexameters, a project he would continue to work on for the next 19 years. After completing the course in Darmstadt he approached, among others such as IBM, the Classics Department at Tübingen University to gauge their interest in his emerging expertise. Though there was no tradition in the Department of applying computing to philological problems they quickly grasped the significance and potential of such approaches. Fortunately, this happened just when the computing center, up to then part of the Institute for Mathematics, was transformed into a central service unit for the university. Drawing on initial funding from the Physics department a position was created for Ott in the Tübingen Computing Center. His role was to pursue his Latin hexameters project and, above all, to provide specialised support for computer applications in the Humanities. In this interview Ott recalls a number of the early projects that he supported such as the concordance to the Vulgate that was undertaken by Bonifatius Fischer, along with the assistance they received from Roberto Busa when it came to lemmatisation. He also talks at length about the context in which his TUSTEP programme came about and its subsequent development. The interview strikes a slightly wistful tone as he recalls the University of Tübingen's embrace of the notion of *universitas scientiarum* in the 1960s and contrasts this with the rather more precarious position of the Humanities in many countries today.

Biography

Wilhelm Ott was born on 3 January 1938 in Gerolzhofen, Germany. From 1949 to 1957 he attended the Altes Gymnasium Würzburg. From 1957 to 1966 he read Philosophy at the Pontificia Universitas Gregoriana and Theology and Classics in

J. Nyhan, A. Flinn, *Computation and the Humanities*, Springer Series on Cultural Computing, DOI 10.1007/978-3-319-20170-2_4

the Universities of Würzburg, Tübingen and München. He was awarded a PhD in New Testament Theology by the University of Würzburg in 1965. He was a research officer (*wissenschaftlicher Angestellter*) for computer applications in the Humanities at the Computing Center of the University of Tübingen from 1966 to 2003 and, from 1970, head of the Division for Literary and Documentary Data Processing, which had been founded for this purpose and where the Tübingen System of Text Processing Programs (TUSTEP) was developed. He also had various other roles: from 1967 to 1970 he acted as the system administrator for the mainframe computer of the computing center, and from 1973 to 2003 vice-Director of the center. He was also engaged in university knowledge transfer and commercialisation from an early stage: in 1973 he was co-founder of the limited liability company *pagina*. In addition to his many other activities he was appointed honorary Professor at the Universities of Würzburg and Tübingen in 1988 and 1989 respectively. He officially retired in 2003 but continues to work in *pagina* and acts as head of the Tübingen group that is tasked with the further development of TUSTEP. In 2007 he was given the Busa award of the Alliance of Digital Humanities Organisations (ADHO) in recognition of his outstanding contributions to Humanities Computing.

Interview

JN In 1966 you had your first contact with computing as a participant of the programming course *Nichtnumerische Datenverarbeitung* (non-numeric data processing) at the *Deutsches Rechenzentrum* (German Computing Center) in Darmstadt. The question that I want to ask goes back to a little bit before then. I wondered about your earliest memory, in any context at all, of encountering computing or computing technology?

WO I do not remember too much regarding computing from this earlier period. In 1966 I noticed an announcement on a notice board at the University of Munich (where I was studying Classics) that the *Deutsches Rechenzentrum*[1] was offering programming courses that were also available to Humanities people. This interested me. I had been busy studying Theology and had completed a doctorate in it. I knew that I would not earn my living from Theology and so I had started a second study in Classics. From school times I always had a great affinity to mathematics and physics. I saw the training that was advertised as a chance to get involved in computing in the context of Humanities.

[1] 'In October 1961 … the German Computing Center was founded in Rheinstraße 75 in Darmstadt. It was the first academic computing center in Germany and was one of the most important milestones in Computer Science in Germany' (translation by Nyhan). See: https://www.informatik.tu-darmstadt.de/de/aktuelles/neuigkeiten/neuigkeiten/artikel/50-jahre-deutsches-rechenzentrum-fraunhofer-sit/

JN And so you attended the course in Darmstadt essentially because you saw the notice, and thought, "that looks interesting"?

WO Yes, I thought it looked interesting and I had also some research problems which could perhaps be solved with it. In Theology I had worked on the New Testament and the differences on the teaching on prayer between the gospels (Ott 1965). In this context it was important to look for details in the speech and wording that the respective evangelist used. A lot of philological tools were available, there was special grammars and concordances for the New Testament – and I missed this in Classics. Therefore I thought that a computer could help.

When I started the programming course in the spring of 1966 it was relatively early times for Humanities Computing. The first part of the coursework was in Assembler, this means working very near to the hardware of the machine. The second part was in FORTRAN, using a set of sub-routines for character and string handling that the Darmstadt group had just developed, because FORTRAN, at that time, did not even have a CHARACTER statement. With those sub-routines one could at least get access to single characters and to strings, and one was able to move strings and to collate strings and so on. After some days I found the exercises they did a bit annoying. Therefore, since I was working on Vergil's *Aeneid*, an epic poem written in dactylic hexameters, and since I had learned from Eduard Norden's Commentary on Book VI of the Aeneid (1957) how important it was to also pay attention to the "pictorial elements" of the hexameter when interpreting the poem, I tried to design a program to automatically compile the metrical characteristics which Norden had collected in the appendix to his commentary. It worked, and it was my first experience of thinking about the application of computers to the Humanities.

When I started in Darmstadt I had just the basic tools. I had a FORTRAN compiler and I had that set of sub-routines. Later, when we moved from Darmstadt to Tübingen, it was an additional effort just to provide a set of sub-routines that were compatible with the Darmstadt ones and that would allow me to continue my work on a Control Data computer. In Darmstadt I had worked on an IBM 7090, and later a 7094.

JN What did you think about the computing that you encountered on that course? How useful or difficult was it? What is your general recollection of your feelings towards it?

WO Well, I think it was challenging and I was very curious to see if it would work. I tried hard and the courses (in assembler language and FORTRAN) lasted for a fortnight each. Of course, it worked, and it was fun, and I was happy.

JN In 1966 you became Research Officer (*Wissenschaftlicher Angestellter*) for computer applications in the Humanities in the Computing Center at the University of Tübingen. I'd especially like to hear about what your job entailed and your recollections of some of the earliest projects that you supported and worked with.

WO Well, the programmes that I wrote for analysing the hexameter were not finished by the time the training course in Darmstadt had finished. I was not able to pay for the computing time after the course. The course itself was free, but then, after a fortnight or so they said, "well, now your work is perhaps a research project and you have to pay for your computing time". Computing time then cost 230 DM per hour, and this was a bit much for a student who lived on a scholarship of 400 DM per month.

Therefore, I went to the Classics Department in Tübingen University where I continued my studies and showed them what I was working on to see whether they were interested. At the same time, the computing center had just moved out of the Mathematics Department and had become a central unit for the university, comparable to the university library. They saw the chance that in these circumstances they could also get advice for Humanities applications because I was a Humanist with a doctorate and I had proven that I had also some knowledge of computing. Therefore, they hired me, and the first thing I had to do was to continue my hexameter project but, of course, to also give advice and make it available to other interested people from the Humanities. With time those projects came.

One of the first projects that came was from outside the university. It was the concordance to the Vulgate, the Latin Bible, by Father Bonifatius Fischer (1977), a great project sponsored by the *Deutsche Forschungsgemeinschaft (DFG)*. They had heard about me via IBM Germany, whom I had also contacted in order to find out what opportunities existed for continuing my work. I had contacted Dr Hübner of IBM in Sindelfingen who had also been a member of the Classics Department in Tübingen before he went to IBM. From this contact resulted the first contact with Bonifatius Fischer, more or less a year later.

Another large project also came from outside Tübingen. Prof. Kurt Aland from the University of Münster wanted to prepare a new critical edition of the New Testament and so he wanted to determine which of the many manuscripts of the New Testament could be omitted from the apparatus. This problem required the grouping of the variant readings found by the manual collation of 98 selected passages of the Catholic letters, so as to find out the relationships between the approximately 500 manuscripts that contained the text.

Other projects came in from the University of Tübingen. In 1969, one of the larger projects was an index to 75 volumes (1895–1970) of the *Theologische Quartalschrift.*[2] Then in 1970, an index to the works of the middle-high German poet Heinrich Kaufringer was to be made by Paul Sappler of the German

[2] 'The *Theologische Quartalsschrift* is edited in Tübingen and is an academic journal that addresses all aspects of the discipline of Theology … it was founded in 1819' (translation by Nyhan). See: http://www.thq-online.de/wir_ueber_uns.php. For a short report on preparing the index volume, see http://www.tustep.uni-tuebingen.de:/prot/prot1.html#kustermann. The index was published in 1975 (Seckler 1975).

Department.[3] It was to be an edition plus index. He had asked for the index and I advised him to also prepare the edition itself by computerised typesetting because then the data would be error-free for the index. In the meantime the TUSTEP typesetting programme (see below) was working. It was one of the first typesetting routines that could make up whole pages on a Digiset (the first digital typesetter). And, at the same time, a further project came from the Institute of Hebrew Studies at the University of Tübingen, an edition of the *Mishna* by Michael Krupp. This meant further problems and further challenges because the Hebrew fonts needed for the Mishna edition were not available for the Digiset at that time. The right to left reading was also difficult for processing and so on, but as a former student of Theology, I knew enough about Hebrew that I could get involved with this project.[4]

JN I want to go back and ask you a little bit more about some of those projects. Can you tell me more about Bonifatius Fischer? I know you put him in contact with Busa regarding lemmatisation.

WO Well, there are many interesting aspects in this project. The first thing I told him when he asked me to work on the concordance was "well, before starting we should have an error-free, machine-readable text of the Vulgate". The edition for which the concordance was planned was the new edition which he had just finished, together with Robert Weber. Fischer told me that he was not very good at proofreading; he was much better at typing. Therefore I told him, "well, then let us transcribe this text twice and compare the two transcriptions", and so we did, and it worked out very fine. I have also published a short article about this approach (Ott 1970).

Then, in order to prepare the concordance, the word forms had to be lemmatised (which means using the word forms not as they occurred in the text but rather their dictionary forms, or *Grundformen*). And so I contacted Father Busa in Gallarate, or in Pisa, where he had just moved to, and asked him if he could give us a copy of his *Lexicon Electronicum Latinum* (LEL), as he called it.[5] I visited him in Pisa and he generously gave us his lexicon and we took it as a basis for the lemmatisation. Later he asked me to give him a list of words which occurred in the Vulgate but not in Thomas of Aquinas so as to complement it – it worked fine.

JN You showed me print outs of a lot of this material yesterday. It was fascinating for me to have the opportunity to look at your archive.

---

[3] Sappler, who died on 14 April 2010, was also active in the wider Humanities Computing community. His obituary is here: https://www.uni-tuebingen.de/en/news/newsletter-uni-tuebingen-aktuell/2010/1/leute/12.html. For a short report on the project see http://www.tustep.uni-tuebingen.de/prot/prot1.html#sappler

[4] For outline details, see http://www.tustep.uni-tuebingen.de/prot/prot2.html#krupp and http://www.tustep.uni-tuebingen.de/prot/prot12.html#krupp.

[5] Busa wrote '[f]or this, we first punched, sequenced and numbered the 90,000 lemmas in the Forcellini's *Lexicon Totius Latinitatis*' (1980, p. 86). A description of the 'LEL de Gallarate' by Busa is available at http://promethee.philo.ulg.ac.be/RISSHpdf/annee1969/02-03/RBusa.pdf.

WO Yes. Regarding the lemmatisation itself, the first step was to isolate all the word forms that occurred in the text. A speciality of this concordance was that it included not only the words of the text but also the entries that appeared in the *apparatus criticus*. So the apparatus entries were also transcribed. The second step was to sort the word forms alphabetically. The third step was then to mix into this sorted material the entries from the Busa lexicon. For each word form of the Latin text of the Bible we got an average of two and a half lemmata, depending, of course, on the frequency of the homograph word forms in Latin. For example, *est*, 'he is' or 'it is', could not be only a form of *esse*, 'to be' but also of *edere*, 'to eat'. These are homographs. Another simple example is *facies*, which can be 'the face' or it can be a verb form of *facio.*

Then, after the lemmata had been intermixed with the sorted list of words, they were sorted back into the sequence of the text because otherwise you cannot determine which form corresponds to which lemmata in a given context. Then, a new printed list was produced again, where after each line of text all the word forms that occurred in this line were printed in separate lines along with the respective lemmata. Then the manual work that established which of the lemmata was appropriate in that place started. For this work Fischer had engaged a monastery of Benedictine nuns in Kempen on the Rhine. All of them had an *Abitur* (final secondary-school examinations) with Latin as a second language, and therefore they were very happy with this work which was of course, closely controlled or surveyed by Bonifaitus Fischer. They did work for some years on the lemmatisation.

For this step of manually controlling the word forms for which more than one lemma had been found in the lexicon, I had arranged the materials (that is the lines showing word forms plus lemma below each bible verse) so that the most probable lemma was the first one listed. For example, for *est*, the first was not *edere* but *esse*, though *edere* is, in alphabetic order, before *esse.* I tried to pay attention to the frequency in order to have the most probable solution in the first place. When, in this list, the first lemma was right, then nothing had to be done. If the second or the third lemma was the correct one then the following had simply to be typed: the number of the line and of the word which was printed in this list plus the current number of the correct lemma. Then, the material was sorted back into alphabetic order again, not according to the inflected forms but according to the lemmata. Then the material was prepared for printing using the typesetting programme that had also been developed in the course of the work on this concordance. It was from this project that I first learned that it was important to have a programme that could transport the results of philological work error-free to print. But that is a totally different theme of course!

At the time the project started, the typesetting industry was still relying on hot metal typesetting, lead typesetting. Indeed, due to a hint from Dr Hübner of IBM, the first contact that I had in this context was with the printing house of the *Mittelbayerische Zeitung* (a newspaper publisher) in Regensburg, who had a Linotype driven by paper tape. This paper tape was being prepared by a computer programme, which just provided for automatic line breaks including correct hyphen-

ation. Well, hyphenation and line division were not important for us because the content of the Vulgate was short enough that it normally fit into a line and hyphenation for Latin was also not a problem for the concordance. The control of the typesetter itself was a problem. I had just written the first test when the notice came that the first cathode ray tube typesetter had been installed in Neu-Isenburg. Then I left the programmes that I had started for the Linotype and looked for a way to get access to the Digiset for the typesetting. Digiset was produced by the Hell company in Kiel and was the first cathode ray tube typesetter (see, for example, Sassoon 1993, p.76, 78, 88); in the States it was marketed as VideoComp by RCA Corporation. I got in contact with Lux Bildstudio in Neu-Isenburg and then I tried to prepare a programme so that this typesetter could be used for the publication of our data.

JN Yesterday you mentioned a conviction that you developed quite early on. It was that you wouldn't apply methods that you yourself didn't understand, or weren't familiar with. Would you again tell me about that and the context in which it came about?

WO Well, such concerns arose in connection with some later projects. The first projects I had supported related to Classics, a research field with which I was more or less acquainted. I knew Theology, Latin and Greek, the Bible and I also knew a bit of textual criticism from my dissertation, where it had been important to know how the text had developed over time. And I had been aware that it's a very, very sensitive field with very sensitive problems. For example, the problems of sorting. Many people do not really know how a text should be sorted. I remember many German publications from the early days where the umlauts, for example, were displaced to the end of the alphabet, after 'z'. Similar problems occur with other languages too. From observing such simple problems one concludes that some projects were overseen by a person who was not acquainted with the problems of the respective field and that the solutions that were offered were just not acceptable. Therefore, I decided to provide tools only for problems from my own field, or via a collaboration with persons who really knew their field and who were available and willing to spend some of their time on such discussions.

JN Is it correct of me to conclude that your 'philosophy' about the role of the computer was that it should support the advancement of humanistic knowledge, as opposed to being something to experiment with, and to purposely break things with …

WO No. At that time, people who decided to use a computer were people who had problems [laughs]. Sometimes I say, "well, it was the time of close reading and not the time of distant reading" [laughs]. It was not about playing with the material that was available: they had a problem and they wanted to solve it or to provide tools like dictionaries or indexes to periodical retrieval tools, and so on. Playing around was sometimes also a partial motivation, but the normal work was just helping people to solve problems.

JN You've already mentioned a number of important people in Germany and internationally. I wanted to ask about the key people who you came into contact with, how quickly you started to come into contact with them and also about processes of the transfer and discovery of knowledge.

WO In addition to the contact with the professors at Tübingen who provided that the computing center was established with a person and facilities for the Humanities, the first important contact I had was Dr Hübner of IBM. It was just 3 months after I had finished the programming course, and I wanted to ask him what facilities existed, and how one could proceed in this field. He had just written the hyphenation of German programme for the Linotype typesetter. He also had many contacts and it was due to him that the contact to Father Fischer had been established. The contact with Father Fischer also came about via the publisher Frommann-Holzboog in Stuttgart. From the course in Darmstadt and also from Hübner I got the names of some people and some projects and I tried to contact them too.

There were also some people outside of Germany. I contacted Professor Louis Delatte of the *Laboratoire d'Analyse Statistique des Langues Anciennes* in Liege[6] relatively early in July 1966. I visited him in October and reported on my approach and he found it very interesting, and gave me the hint that it should be advisable to publish or to prepare a paper on it, and to publish it in his *Revue*. So, in fact, in the last number of the *Revue* of 1966, I had the first paper on my computer-aided hexameter studies (Ott 1966). Many people worldwide, who then contacted me, found this paper interesting. One of the first was Joseph Raben from New York, who wanted a notice for CHum. I don't remember the first contact that I had with Stephen Waite (of Dartmouth College), who had been editing *Calculi*, his bi-monthly periodical for computer applications in the Classics since 1967. Father Busa, I just mentioned and reported that I contacted him in the context of the concordance to the Vulgate in April 1967. I visited him and Antonio Zampolli (his assistant at that time) in July of 1967. The contact to Kurt Aland from Münster had also been established by Bonifatius Fischer, who as an editor of Vulgate, of course, had close contacts with him.

Other international contacts came via the hexameter project. In September 1969, there was a large conference of *La Fédération internationale des associations d'études classiques* (FIEC) in Bonn.[7] It had about 800 participants, and they also asked me to give a report on my hexameter project (Ott 1969), and on this occasion,

[6] 'Founded at the University of Liège in November 1961, the Laboratory for the Statistical Analysis of Ancient Language (LASLA) was the first research centre to study the classical languages Greek and Latin using computational techniques' (translation by Nyhan). See http://www.cipl.ulg.ac.be/Lasla/

[7] *La Fédération internationale des associations d'études classiques* (FIEC)/The International Federation of Associations of Classical Studies (FIEC) 'is an umbrella organization that covers most associations of classical studies of national importance around the world … Every 5 years, FIEC holds an International Congress which gathers scholars from all parts of the world and from all sub-fields of classical studies'. See http://www.fiecnet.org/#!mission/ceax.

I came into contact with, for example, David Packard, who later founded the Packard Foundation.

Other contacts came too. Just about 4 years after my appointment to the Computing Center, in 1970, I heard of a conference Professor Roy Wisbey was organising in Cambridge. I wrote him a letter to ask if I was allowed to come to this conference, which seemed, at first glance, to be a national conference for England. In fact, I was the only German participant there and through it I made contacts with other important people, for example, Susan Hockey.

JN How did you find the reception, and the sense of cooperation, or not, among the participants?

WO I did not have a paper there because I noticed it relatively late and I was just in the audience. But Roy Wisbey replied and invited me and it was a very fruitful contact. I am convinced that my presence there was also important because when the ALLC was subsequently founded it was not as a national British institution but as an international one. This is also what Wisbey told me later.

JN I wanted to ask you a little bit about the founding of the ALLC, so seeing as you've mentioned it, shall we talk about that now and then I might again step back in the chronology.

WO We can, of course. There was the session in 1973, I think, when I was not present but they had asked me before then to be the German representative in the Association. I decided not to do so and I was also a bit late with my answer. So, Professor Lenders of Bonn[8] was proposed for this and they asked me to be the representative for a specialist group on textual editing techniques. I heard the details of the founding second hand because I was not present at the founding session. There was a second conference in Edinburgh in 1972, before the founding of the ALLC and after the Cambridge conference where I was also present and gave a paper (see Ott 1973).

JN So, I want to go back and ask about the metrical analysis that you published with Niemeyer, between 1970 and 1985 (see, for example, Ott 1970). Yesterday you showed me, what I might call a 'paper search engine' [both laugh] to the text made with punched cards. Will you please describe them and explain how they worked?

WO The problems that I wanted to solve (in addition to providing overviews for the hexameter poetry) I had drawn from the appendix to the commentary of Eduard Norden to the sixth book of the Aeneid (1957). He was convinced that metrics were important for interpreting a poem and had a lot of criteria that he looked for: the

[8] Winfried Lenders (1943–2015) was appointed Professor for 'Linguistische Datenverarbeitung' (Linguistic data processing) at the University of Bonn in 1974. A short obituary can be found here: www.gscl.org/ehrenmitglied.html.

number of words and the position of the word endings in respect to the verse structure. In the middle of a hexameter there is normally also a caesura (or a pause) and he also looked for where exactly this caesura is on average, or in most verses, and so on.

Therefore, one of the tools I provided, and which I thought it was possible to provide beyond the printed lists, was a tool to allow one to look for combinations of word endings in the verse. I thought that it could be done relatively easily using a punched card. The punched card had 80 columns, with at least 10 positions that could easily be numbered vertically. Additional rows 12 and 11, as they were called, were not used for representing the number of lines, or the number of verses. Therefore I provided 16 punched cards, one for each position in the hexameter, as the hexameter consists of 6 feet, and each foot can have either two or three syllables: two long syllables, or one long syllable and two short syllables (that makes 16 times three, or 18, but the verse end is always a word end. Therefore it can be neglected and the last foot is almost always two syllables only. That meant I had 16 positions that were interesting). And so, I provided 16 punched cards. On each card I made a hole in the respective position. Where, for example, a word ended just after the first syllable in line three of a poem, I made a hole in the first card in column zero, row three and this indicated the occurrence of a monosyllabic word at the beginning of the verse. And this I did for the 16 positions in the verse and for every line. Then, if you want to see if, for example, a verse that starts with a monosyllable, and ends with a monosyllable, you just take the first and last card and put them together, one above the other, hold them against the light, and where the holes are shining through, there you have the number of the lines of the verses which start and end with a monosyllabic word. It's as easy as this.

JN And where did the idea for this come from?

WO Well, I was accustomed to punched cards. Data entry was on punched cards and some output was on punched cards for further processing. The compiled programmes were also on punched cards. So, for a second run, if you have the same programme but different data, you could just use the binary text of the programme to produce it. I was also aware of some people's work with so called *Randlochkarten* (edge-notched cards) where one could sort the material by mechanical means.

JN *Randloch* is the hole at the side of the card?

WO It was cards where the content was written by hand. On the margin of those cards was a perforation, I think it was, and you could cut this with the help of a special scissors, so that if you got a needle or a nail or something to go through a notched hole and lifted the needle, the respective cards would fall back. This is a mechanical tool and I thought such approaches to inspection could aid this problem.

JN Wonderful. And do you know of any other Humanities Computing projects that did that?

WO No, I don't remember any at the moment.

JN In 1973 you co-founded *pagina GmbH,*[9] so I'd like to hear about how it was that the idea for a commercial company came about?

WO Well, it also has a long history. I related that I came into contact with the Digiset typesetting technique in the course of the work on the Vulgate. In fact, the first volume that I published with a programme that I wrote for the Digiset was not the Vulgate, but the first volume of my hexameter studies. One of the other earliest projects to use this typesetting program was the edition of the works of Kaufringer by Paul Sappler, which was to be published by Niemeyer. Paul Sappler asked Niemeyer if, instead of delivering a manuscript, he could just typeset it. The technical production manager of Niemeyer, Wolfgang Reiner, saw it just as he was looking for a replacement for hot metal typesetting. He got in contact with me and proposed to make this service available commercially for typesetting in publishing houses. We decided to found the firm *pagina* (*pagina* is a Latin word for the page). The programme I had made was the first programme, as far as I know, that did not only provide hyphenation and line breaks but also complete page make-up, including page numbering, running heads and so on. So we founded *pagina:* the name comes from the ready-made pages we created, and it was a typesetting firm for, more or less, publishers in the Humanities, because Niemeyer was, of course, also in this field.

JN And *pagina* is, I think, probably unusual among Humanities Computing and Humanities projects in that it is an early example of research that was done in a university and then commercialised and taken outside the university context. And so I wondered what sort of a response you got to it from your university colleagues?

WO Well, it was not very common at those times, to get a so-called *Ausgründung* (spin off), as they call it nowadays in German. Well, I, of course, made applications and it consisted more or less in the fact that I was allowed to get *Anteile* (financial shares) of a *GmBH (Gesellschaft mit beschränkter Haftung;* a limited liability company). The cooperation with *pagina* was indeed really fruitful for the university and its publishing because I was not trained in publishing or in typography and so on. I got this whole know-how from this company and from Wolfgang Reiner. Also the first description of the typesetting programme was not made by me but by Reiner as he also knew the terminology and other things that were important. So this was a really fruitful cooperation.

JN And we should make the point that *pagina* is still going strong to this day.

[9] See: http://www.pagina-online.de/.

WO Well, *pagina* has become an established firm in the meantime. At the moment typesetting is only a tiny part of its activities. This is another theme. The first business area of *pagina,* used as a supplement to the name of the firm, was *Elektronische Satzherstellung* (electronic typesetting) and later, in 1966, it was changed to *Gesamtherstellung Wissenschaftlicher Werke* (overall production of scholarly works).

Nowadays (and since 2011) it is "publishing technologies", including online publications and e-book publications, workflows for publishing houses, the introduction of XML technologies and so on. Typesetting is only a tiny part of it, but *pagina* in the meantime has also established an *Abteilung* (a division) called Digital Humanities, and it's also giving advice to Humanities Computing projects from transcription to collation and so on. At the moment two to three people are working there.

JN Now I want to ask you a bit about TUSTEP (*Tübinger System von Textverarbeitungs-Programmen*; a professional toolbox for the scholarly processing of textual data).[10] So, from what I've read, it was named TUSTEP in 1978, but it seems quite clear from what you've been saying, that the processes towards it were going on for some time before that. You've explained a lot of the context that TUSTEP came about in but I also want to ask about the process of identifying that TUSTEP was needed and useful and how you went about those first steps of actually setting it up?

WO As I said, I was hired by the university to give advice and support to Humanities projects. The first thing I did was to make available the tools I used (this was the set of sub-routines for character handling in FORTRAN) and to give FORTRAN courses. Paul Sappler, and Gottfried Reeg (from 1984 at FU Berlin) started with those tools and programmed by themselves. We did the programming for other people in the computing center. As the number of projects increased, it was no longer possible to continue as before. Therefore we thought of a way to make the users a bit more independent by allowing them to do much of the work for themselves. We tried to isolate the most common basic functions that were needed for Humanities Computing.[11]

One, for example, was a programme, it was indeed a separate programme in the beginning, which extracted particular sentences or records or a certain string from a file, for example. Another programme, it was also a separate programme, did search and replace functions on a file. Other basic functions related to preparing an index and allowed one to firstly break down the text into index entries, then to provide a sort key for the index, then to sort the material and then to reduce the sorted material to index entries. For example, I don't want to have 750 individual entries for the

[10] See: http://www.tustep.uni-tuebingen.de/tustep_eng.html.

[11] In this context it is interesting to note the later work of Unsworth (2000) on 'scholarly primitives' or methods that Humanities researchers are thought to have in common.

copula *und* (and) along with 750 individual references to it. I want the information to be summarised in one line. Neither do I want 750 references to frequent words, I only want their frequency, for example. Proceeding in such a modular way was one of the lessons I had learned through the concordance to the Vulgate.

When we started the concordance there were just two or three programmes available that produced concordances. One of them was COCOA, the word count and concordance programme provided by ATLAS in Great Britain, which later was the basis for the Oxford Concordance Program (OCP) by Susan Hockey and Ian Marriot. Susan Hockey was at Atlas before she moved to Oxford and co-developed the OCP programme. There was a further programme in Regensburg COBAPH (COBOL basic programmes for Philology), which also produced concordances, but these programs were more or less black boxes. You had some parameters that you could give along with your input text, then you got a concordance, but it was not lemmatised.

So I thought that, as problems are different in every field of the Humanities, we should provide the basic functions. I just mentioned that we included the critical apparatus in the concordance to the Vulgate. Therefore, I realised that the procedure for decomposing texts must be flexible and that it should be possible to define the details by parameters given by the user. The second step was providing the sort key. If you sort an English text or a Latin text, an ASCII key is just fine. However, when dealing with German texts, for example, you must take account of two different rules. One is for sorting lists of proper names, where you sort the umlauts as 'a + e' or 'o + e' or 'u + e', whereas you have to sort the umlauts as the basic letters without the trema or the umlaut dots in all other cases, as is done in subject indexes or in a dictionary. I think that if people from different faculties and different subject areas are to be helped, then we must provide the basic tools and the ability for them to define them according to their needs. Then the sorting (if the sort key is ok) is a purely mechanical thing. So this is also a separate module, which provides some efficiency and so on. The third step in this process of producing an index is, as I just said, to prepare the entries that have been sorted in the form required for publishing in print or on the web etc. If, like in the concordance to the Vulgate, you have a normal wordlist, it's relatively easy. If the text is indexed according to the subjects it contains, you will have a hierarchy of subjects, with a heading concept and subconcepts and you will also want to show this. In such a case you should have also the possibility to define in detail how these records are to be built and presented.

With such elementary modules, users no longer had to care about programming in FORTRAN (or a similar language). They had those modules and they could specify input files and parameters, and they got an output file which could be used as input for the next module. Well, this is the concept we had; and in the course of time the modules turn out to be rich and sufficiently complete, so that in 1977, I think it was, I gave my last FORTRAN course. From 1977 we just instructed users on using those modules, and in 1978, when the child had some maturity, we baptised it, and called it Tübingen System of Text Processing Programmes, TUSTEP.

JN From 1973 until 2004 you organised the *Kolloquium zur Anwendung der EDV in den Geisteswisschenschaften an der Universität Tübingen.*[12] So when you look back at this very successful and, I think, symbolic symposium series, what are the real highlights for you, in terms of the papers given and those who attended?

WO We founded the colloquia to offer current users and interested guests the possibility to get together with us and with each other in order to share problems and to learn about what was happening outside of Tübingen. The colloquia started in 1973 and this was also the time when ALLC was founded. There is some correspondence in the dates because I wanted to keep in contact with international developments. In the first colloquia we had only relatively short papers that gave an overview of what was going on. Later, we tried to begin with a short overview of news followed by two papers (per colloquium) that got more into the details of the problems at hand. I invited people from outside from relatively early on. They were not using our system (TUSTEP), but other systems and this allowed me to learn what was happening elsewhere, which methods were applied in other places and so on. And so, with time, I had the chance to invite important people from the international community.

One of the most prominent speakers was Father Busa, in 1990, exactly 30 years – to the day – after he co-opened the colloquium on '*Maschinelle Methoden der literarischen Analyse und der Lexikographie*' in Tübingen in 1960. The last speaker on 5 February 2005 was John Unsworth on the 'Importance of digitisation and cyberinfrastructure in the Humanities'. Other speakers included Harold Short, on 18th of November, 2000 on 'The Role of Humanities computing: experiences and challenges'. I had planned to invite Antonio Zampolli to that colloquium, and he had just also consented to come, but then he had an accident, and could not come and Harold Short sprang in more or less immediately.

JN And when you think of the content of the papers given, is there one that really stands out in your mind as having been very exciting or one that heralded a new step forward?

WO It's difficult to say. The papers we had from the international participants were mostly overviews of what was happening. I could not say that there was one or two or three that were exceptionally important. People were coming from many projects across the whole field of Humanities Computing. This not only included texts but also archaeology, not very often, but sometimes, and so on. Well, I could not really say what was exceptionally important. Detailed reports of larger projects, for example, the Leibniz edition by Professor Schepers from Münster influenced other peo-

[12] See: http://www.tustep.uni-tuebingen.de/kolloq.html.

ple in their work. I'm just going reading over the list of participants. Susan Hockey was here in December 1998; David Seaman in February 1999; Michael Sperberg-McQueen in 1995; Jean-Louis Lebrave in 1986; Johann Cook from Stellenbosch in 1986; Michael Krupp from Jerusalem in 1977 …

JN You've already mentioned Eduard Norden. I wondered if there are any other people, ideas or books that influenced your approaches that you'd like to mention?

WO Eduard Norden was for the hexameter project, my own project. The other influences on the development of the software came from the particular projects. I do not remember, at the moment, a publication that had so much influence on me. Mostly the problems came from the projects, and sometimes of course, from publications I had read, but most of this was in collaboration with the project leaders. From the beginning I was also visiting the respective conferences and heard the papers that were given there. I had visited all of the early ALLC conferences and when ALLC merged with ACH I decided to go there every second year only. Much *Anregungen* (stimulation) came out of those conferences, of course. I think it was really important that for a position like the one I had, at a central institution like the computing center, that the person who is there is not only fit in Informatics (the concept of Informatics wasn't available at that time) but that they also know the methods and techniques of research in the Humanities and is at home there. I always said, when you have a technological problem it is relatively easy to get acquainted with the respective informatics methods and technology. That is much easier than it is for an Informatics scholar who must learn what Humanities problems are and the details that should be researched.

JN Yesterday I was referring to the fact that you had not one but two honorary professorships and I asked if you would have liked to have received them earlier. You said, no, because you had such freedom in your position in the computing center. Can you tell me more about that because I think it's very interesting from the international perspective?

WO Well, the first honorary professorship I got was in Würzburg, at the university, because they had established a course for Humanities Computing, and they wanted to have a Professor on their list. They selected me to be this person, and they gave me this title, and only shortly afterwards, a year or so later, the University of Tübingen gave me this title in order to give more weight to a project we had started in Tübingen. For the Universities of Baden-Württemberg there had been created *Forschungsschwerpunkte*, or centers of research expertise for a certain field. In Tübingen we were leading in *Wissenschaftliche Textdatenverarbeitung* (scholarly text data processing) and this was the context in which I got the second professorship. In Würzburg it was for the teaching aspect, but in Tübingen it was for research and responsibility for the project.

Had I been a normal professor, I would not, at least nowadays, have had the freedom that I had to develop those things. It would have entailed too much administrative work. Well, there was a lot of administrative work with the other position I had, especially since I was also the vice-Director of the computing center. All of this sometimes entailed a great deal of work, of course, especially when, in 1988, the Director changed in the middle of the year and I had the whole burden on my shoulders. But otherwise I had all the possibilities I wanted and I also had up to four collaborators who worked for my Department and I was very happy in these things.

JN You also mentioned that it was originally the Physics department that had some funds for the Computing Center?

WO When I started in Tübingen in 1966 the Professors of Classical Philology (Prof Ernst Zinn and Prof Wolfgang Schadewaldt) asked the computing center to create a post (that I would fill) in order to give support to the Humanities. The Physics Department had a post that was then free and they lent this post to the computing center for a year. It would not have been possible some years later. Well, it was in the late 1960s and I've always said that it was a time when the university was still taking account of the meaning of *universitas scientiarum* and individual disciplines had more or less equivalent weight. Well, when one sees what presently is happening to the Humanities worldwide and how they are going down in support one realises that it would not be possible today.

JN This question is probably not so relevant to you because the people who would have come to you were interested in using computers in any case. Still, I wanted to ask about your impressions of scholars who were not using computing in their research and about their evaluation of Humanities Computing research?

WO Well, I had little contact with those people. I remember one, whose name I will not mention here [laughs] because he's known, who, in a relatively important session said that the use of computers in the Humanities should be forbidden. But this statement was not very well accepted! As for myself, as I told you, I came to the Classics Department here in Tübingen, which did not use computers or even know that computers could be of help to them and I was accepted with open arms. It was seldom that I had contact with persons who were hostile to computing. There were some exceptions, of course, the one I told you about was in the context of the German Research Association.

JN Looking at the development of the field up to the present, do you have any disappointments about routes that the field took or didn't take?

WO Well, it's a question that I perhaps cannot really answer. The field that we tried to support was the field where Humanities people had problems that were not necessarily the mainstream of the field. Preparing an edition is a very laborious task and nobody becomes really known for having prepared a good edition. Indexing is also hard work and nobody becomes famous for preparing a fine index. One would perhaps say that those times and problems were connected with close reading and not with what nowadays is called big data and distant reading. Of course, the development of the web and the availability of material, and so forth, requires those questions to be asked. Yet, sometimes I have the impression that they are asked with too little real knowledge of what is being asked. You and I have spoken about the case of N-grams, where the results were not as exact as they should be because, for example, they could not consistently differentiate between the long s and f in some scanned documents. Nevertheless, big data and the availability of the data opens the possibility of asking new questions which could not otherwise be asked about the development of language, and so on. But it's another field of problems from that which we tackled.

Our TUSTEP software is more or less a set of tools for single works and it is also for development and for the analysis, editing and indexing of textual documents. It is not only for textual documents, but also for documentation. Many library catalogues have also been made with the tools we provided. At the University of Tübingen the Incunabula catalogue is still based on a TUSTEP server in the background.

JN It was really fascinating for me yesterday when you showed me so many of the beautiful editions that have been prepared with TUSTEP. It is very rare for me to see something material that has come out of these processes.

WO Well, sometimes you do not even notice that it has been prepared like this. Some people did not even mention that it had been done. This was perhaps caused by the fear that some people had in the beginning about applying computing but today it's no problem of course, no question at all.

JN Is there anything that you want to add or anything that you want to bring up?

WO Some people are supposedly against TUSTEP because it is one of the oldest programmes. So, I sometimes say, 'Well, ok, Daimler, Benz, Mercedes, they have been the first to build cars and therefore they are so antiquated!' And well, sometimes this helps as they begin to realise that TUSTEP has not been left in the state of 1966 or 1978; it has been developed. For example, the TUSTEP typesetting programme is the first one I know of that is able to typeset documents with XML encoding and to provide a stylesheet for typesetting them. There are, of course,

some other reasons that TUSTEP is not very well known internationally and, perhaps also in the meantime, less accepted in the German-speaking environment. One factor is the language of documentation: it's in German only. In 1989 we had an English translation of the manual done but this is now out of date and we did not have the means to continue this work because all the funding of the *Forschungschwerpunkt* (mentioned above) had ceased in 1989. The second factor is the command-line-based interface.

Therefore, for some years we have been working on an interface to the TUSTEP programmes in an XML environment. It's called TXSTEP, we just changed the 'u' in TUSTEP to an 'x'. In the meantime we have one user for it, working on the *Faust* edition in Frankfurt. They are using it especially for collating the sources of the *Faust* material. This XML interface for TUSTEP has the advantage that it's based on an XML schema that remains in the background but informs you about what steps are possible, what basic modules are available and so on. Annotations and instructions in pop-up windows indicate in English whether something is possible and this helps a bit to get rid of the language problem and of the necessity of studying the manual in advance. So I hope that such developments will perhaps help to make those tools a bit more attractive, also in the non-German speaking environment and in an environment where you are no longer willing to use command-line-based interfaces. Also, XML is not as 'user-friendly' as one might expect from other computing applications. But at least people who are accustomed to using computers nowadays are used to an XML environment for tools. This will also perhaps help with the switch to an application of this kind.

JN Thank you very much.

References

Busa, R. (1980). The annals of humanities computing: The index thomisticus. *Computers and the Humanities, 14*(2), 83–90.

Fischer, B. (Ed.). (1977). *Novae concordantiae Biblorum Sacrorum iuxta vulgatam versionem critice editam* (Vol. 5). Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Norden, E. (1957). *P. Vergilius Maro, Aeneis Buch VI* (4th ed.). Leipzig: Teubner.

Ott, W. (1965). *Gebet und Heil. Die Bedeutung der Gebetsparänese in der lukanischen Theologie* (= Studien zum Alten und Neuen Testament, Bd. 12). München: Kösel.

Ott, W. (1966). Metrical analysis of Latin hexameter by computer. *Revue (LASLA, Liège), 4*, 7–20.

Ott, W. (1969). *Metrische Analyse lateinischer Hexameter mit Hilfe der elektronischen Datenverarbeitung*. FIEC bonn [conference presentation].

Ott, W. (1970). *Metrische Analysen zur Ars Poetica des Horaz* (= Göppinger Akademische Beiträge Nr. 6). Göppingen: Kümmerle 1970. 122 pp.

Ott, W. (1973). Computer applications in textual criticism. In A. J. Aitken et al. (Eds.), *The computer and literary studies* (pp. 199–223). Edinburgh: University of Edinburgh Press.

Sassoon, R. (1993). *Computers and typography*. Oxford: Intellect Books.

Seckler, M. (1975). *Register zur Theologischen Quartalschrift, Tübingen, 1895–1970*. Mainz: Matthias-Grünewald-Verlag.

Unsworth, J. (2000). *Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this?* King's College, London symposium: Humanities computing: formal methods, experimental practice. http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html