

Chapter 9

There Had to Be a Better Way: John Nitti and Julianne Nyhan

Abstract This oral history conversation was carried out via Skype on 17 October 2013 at 18:00 GMT. Nitti was provided with the core questions in advance of the interview. He recalls that his first encounter with computing came about when a fellow PhD student asked him to visit the campus computing facility of the University of Wisconsin-Madison, where a new concordancing programme had recently been made available via the campus mainframe, the UNIVAC. He found the computing that he encountered there rather primitive: input was in uppercase letters only and via a keypunch machine. Nevertheless, the possibility of using computing in research stuck with him and when his mentor Professor Lloyd Kasten agreed that the Old Spanish Dictionary project should be computerised, Nitti set to work. He won his first significant NEH grant c.1972; up to that point (and, where necessary, continuing for some years after) Kasten cheerfully financed out of his own pocket some of the technology that Nitti adapted to the project. In this interview Nitti gives a fascinating insight into his dissatisfaction with both the state and provision of the computing that he encountered, especially during the 1970s and early 1980s. He describes how he circumvented such problems not only via his innovative use of technology but also through the many collaborations he developed with the commercial and professional sectors. As well as describing how he and Kasten set up the Hispanic Seminary of Medieval Studies he also mentions less formal processes of knowledge dissemination, for example, his so-called lecture ‘roadshow’ in the USA and Canada where he demonstrated the technologies used on the dictionary project to colleagues in other universities.

Biography

John Nitti was born in 1943 in Yonkers, NY. He has been Emeritus Professor of Spanish and Portuguese at the University of Wisconsin-Madison since 2001. His PhD thesis was based on the Aragonese Book of Marco Polo and his degree was awarded by the University of Wisconsin-Madison in 1972. Shortly thereafter he took up the post of Assistant Professor of Spanish and Portuguese and became full Professor in 1985. In the early 1970s, Nitti, with the support of his mentor Professor Lloyd Kasten, began to explore the application of computing to Old Spanish lexicography. He subsequently won major funding from bodies such as the NEH and, in

later years, the government of Spain. Such grants allowed him to devise innovative ways to bring computing to bear on the Old Spanish Dictionary project, which the Wisconsin Seminary of Medieval Spanish Studies had been at work on since the 1930s. In addition to his Directorship of the Old Spanish Dictionary project (1975–2001), in 1975 he and Kasten founded the Hispanic Seminary of Medieval Studies (HSMS), initially to disseminate the large numbers of transcripts and other materials that the dictionary project was producing. Today, this not-for-profit organisation (now based at the Hispanic Society of America in New York) has become an important publisher of texts in Hispanomedievalism and related fields. His many publications include (together with Lloyd Kasten) *The Electronic Texts and Concordances of Medieval Navarro-Aragonese Manuscripts* (1997) and *Diccionario de la prosa castellana del Rey Alfonso X* (2002)

Interview

JN My first question is: what are your earliest memories of encountering computing technology?

John It is a very vivid memory because it really was the pivotal point in my approaching lexicography using as much information technology as possible and, of course, we're talking about a long time ago. We're talking about the very late 1960s and early 1970s. My first significant grant was, if I remember correctly, for \$242,000, from the NEH. I was still a graduate student at the time I developed that proposal. But prior to that time – it's an odd sort of a thing because I know one of your questions here is “which people particularly influenced you and how” (with regard to the technology aspect, obviously not Medieval Hispanism, which is my field) and the two are linked – my first contact with computer technology, with an eye toward employing it and applying it to my research, occurred via one of my fellow graduate students. Actually, he was not one of our best graduate students, but I owe him this. One day he said to me “you know I'm looking for a thesis topic? What I really want to do is to generate a concordance of a large Medieval Spanish text, based on a transcription that I'll do.” At the time, our campus academic computing facility was mainframe-based, as most if not all were. We had mini-computers, of course, but there were no microcomputers at that time. He said “I heard that our campus computing facility bought a concordancing programme and brought it up on the UNIVAC (which was the big campus mainframe at the University of Wisconsin at the time). You're interested in things technical, would you do me a favour and check it out?” I said “OK”.

So I went over to the computer centre and started talking to the people and they gave me a few dollars' worth of credit so I could actually run a sample concordance. Now we're talking really primitive stuff, I mean it printed the concordance but these were the days when you were lucky if your campus had an IBM system for aca-

ademic research, because then you could use extended ASCII and you could get upper and lowercase characters. But if you didn't, you were simply lost because all it could produce and print were uppercase characters and, of course, the input device, was a big disappointment to me as well. The University of Wisconsin-Madison had a very large campus computing facility. It is a campus of some size, we have traditionally had 40,000–45,000 students at this campus from year to year and the University of Wisconsin-Madison is, in fact, the flagship campus of the University of Wisconsin University system. Well, the input device that they pointed to was a keypunch machine, and I'm thinking "What? I have to use punch cards to do this stuff?" Of course they only had uppercase letters. That was it, so you had to punch all these cards and transcribe all these texts in uppercase. Well, I took my colleague (who I mentioned above) over there and I got him started. Eventually he did produce a concordance and he submitted that with an introduction as his doctoral dissertation (we were both working on our dissertations simultaneously).

But at the time I was doing other stuff. I was already working as a graduate assistant for my mentor Professor Lloyd Kasten (see, for example, Jover 2002) while completing my doctoral dissertation. Mr Kasten and I forged a partnership which lasted until his death at the age of 94, though, it is important to note that he was still going strong and still working on the *Dictionary of the Old Spanish Language* (hereafter DOSL) with me 3 months prior to his death. I remember well that when I first proposed computerising DOSL, he took to the idea right up front. At that time he was already in his late 70s and he said "Let's do this. Wouldn't you like to get into this computer thing more deeply and see if we can't computerise the Old Spanish Dictionary project?" I said "I sure would, except we don't have any money to do it!" He said "Well, we'll start out with some."

He was a very frugal man, a single man. He had a sizable savings account so he was totally willing to put up what for an individual were quite large sums of money so that I could start playing around with computers. So it was probably Mr Kasten who was more influential in giving me an opportunity to get my feet wet, or get both our feet wet, as it turned out. But I have to give credit to that one graduate student. That poor guy died young, he became a professor at the University of Wisconsin, Steven's Point, a smaller campus, and he died in his 40s, which was a tragedy. But in any case that's how I got started in all of this.

JN When you headed over to the computer centre and got some credit, were you one of the few Humanities people who had turned up there?

John There was a Professor of English who in fact went before me in this historical sequence and I can't remember his name. Quite frankly, he and I didn't have much contact at all. He was involved with the computational aspects of the *Dictionary of Old English* project under the directorship of Professor Angus Cameron at the University of Toronto, and was therefore into the application of computer technology to humanistic research, such as it was, sooner and more deeply than me. He eventually left the University of Wisconsin for another position and we had no further contact.

As the years passed I got interested not only in developing ad hoc software to do what I wanted for the creation of DOSL but, more importantly, to develop, believe this or not, novel hardware applications. Not ourselves physically, although I did end up doing some microcomputer kit building. These were the early years before there was much in the way of microcomputers: Radio Shack had not yet released the TRS-80, Apple was on the verge of releasing its first Apple, which wasn't very powerful at all and couldn't even do what I wanted to do and IBM had not yet released its PC. In any case, I started looking right from the very outset.

We were obliged to use the campus mainframe because it was in the university's interests since they were renting it from UNIVAC at great expense. They wanted to generate as many users as possible and, up to that point, and even afterwards, the lion's share of users were Scientific as opposed to Humanities people. Things changed a little bit as the years went on. But what happened with the DOSL project was once we got some substantive funding – although Mr Kasten continued to contribute out of his own pocket so that I could indulge my whims testing out devices and I'll mention some to you in a bit – it became patently clear to me that our having to use the campus UNIVAC of the University of Wisconsin for our processing was in fact what we call today 'a rip off'. It was outrageously priced. I'll give you an example. A project such as ours, which presupposed that we were going to be transcribing from the original manuscripts (or photographic reproductions of the original manuscripts), required a number of things. Number one was that I develop a manual of transcription for the *Dictionary of the Old Spanish Language*, an encoding text, if you will. I was given to understand that I was one of the earliest ones around to actually develop such a thing.

In order to publish it, and anticipating the publication of large quantities of data, Mr Kasten and I created a non-profit publishing house called the Hispanic Seminary of Medieval Studies.¹ Here I'll mention John O'Neill, my Irish student. The then Director of the Hispanic Society of America, Theodore S. Beardsley (who had held that post for some 40 years), was a very close personal friend of mine and he asked me "have you got a good PhD who could be my replacement as departing Curator of Rare Books and Manuscripts?" I said "I sure do, he's an Irish man, if you don't discriminate against the Irish!" Ted just chuckled (parenthetically, Ted just died a few months ago). He said "well, I'm coming to Madison". He wanted to meet this Irish boy. He came and met him and liked him as much as I liked him, so he hired him on the spot to be the Curator of Rare Books and Manuscripts. This was some 15 years ago and John O'Neill is still there as Curator but they've made him, in addition to the Curator of Manuscripts and Rare Books, also the Head Librarian of the Hispanic Society of America.

Mr Kasten and I decided that we wanted to begin publishing as soon as possible, even in intermediate forms, the large numbers of texts that we were transcribing from the original manuscripts or from photocopies of the originals. We wanted to distribute them to the world. So, we looked around for another technology: what would it be that we could marry to the fact that we were capturing all these key-

¹ See <http://www.hispanicseminary.org/index-en.htm>

strokes in a computer-readable form? The technology for mass dissemination of data was – there weren't any DVDs or CDs then – 'computer output on microform' (or microfiche in this case). We miniaturised all these thousands of pages of transcription and/or concordances that we were generating because we weren't going to be able to publish this stuff, which was fairly esoteric, in standard book form. There are relatively few people in the world who are interested and willing to pay the fortune you're going to have to charge just to recoup the expenses of all this material. So we said we had to distribute this in a medium that we could afford to distribute for pennies, literally, pennies on a dollar. And what was that medium? It was computer-output on microfiche. So we worked a contract with a commercial service bureau in Minneapolis, Minnesota, and I went up there a number of times and we got it all coordinated and going.

For years we published and disseminated the data that we were generating, both textual transcriptions and corresponding concordances with frequency counts and all that typical stuff that you get with concordancing schemes. And we were able to sell it through this new publishing house we created, the Hispanic Seminary of Medieval Studies. You may want to go online to the Hispanic Seminary of Medieval Studies. John O'Neill in New York has created its website and he actually sells the publications out of that website. John has kept the Hispanic Seminary going.

The medieval seal on the website was drawn freehand by my sister. It's inspired by seals of the thirteenth century for the Kingdom of Alfonso the Wise, who is, in fact, the King for whose original manuscripts we generated the largest database. Many of those manuscripts that Alfonso had produced for him, and that we assume he held in his own hands have survived, believe it or not. And we were able to use photostatic reproductions or microfilm of those manuscripts to create the *Dictionary of the Castilian Prose of Alfonso X*. John O'Neill is keeping that publishing operation going on his own time, he operates it out of the Hispanic Society of America now. When I retired I had it all legally transferred to him and the Society, knowing that he would keep it going.

JN Do I understand correctly that it was the time you spent in the computer centre that inspired you to involve computational technology in your research?

John Yes, but I quickly learned that the state of that technology as applicable to our research was deplorable; quite frankly, it was primitive. I thought "God, we have to print anything we publish out in uppercase letters? That's ridiculous!" All they had were these high speed chain printers (the 15 in. wide paper with the perforations on the side) that fed the paper through these machines at breakneck speed. It was just crap and if they hadn't changed the ribbon, you were bound to get something that could barely be read.

JN And how, despite those limitations, were you able to foresee ...?

John Well, here's how. Initially, I was working hot and heavy on trying to find a data input mechanism – this is hardware we're talking about now – that would have

upper and lowercase letters from the get-go. On the mainframe computer in those days the typical editor you had was referred to as a ‘line editor’. Imagine this: you’re working with a terminal, not with a CRT terminal but with some sort of a teletype terminal, printing out this junk at ten characters per second, or whatever it was that they eventually cranked it up to, it wasn’t much. What you had to do was locate the line that you wanted to edit, bring it up and print it out. Then you’d have to use search and replace algorithms with the editor (I’m talking about an online editor to the mainframe. To make a change you had to go “find this, change to that”). Now imagine how you do that a zillion times to correct the transcriptions which were input. That’s another very important point with my research, the need to get a much better method for inputting data or capturing keystrokes, as we used to say.

I decided early on that I wanted to get away from the keypunch as soon as I could. So, I used as a pilot, basically, the transcription of the very first manuscript of Alfonso the Wise, which was a very large manuscript. We used the keypunch for that and we managed to produce the entire transcription all in lower and uppercase, but we were flagging. I decided we would flag the uppercase letters with an extra symbol so that we’d at least be able to convert all those when we got better technology. In fact this did take place but there was a disaster. We had something like 15 boxes of keypunch cards, in long-ish boxes as they had in those days. We were trying to transport them to the computer centre so we got one of these wheeled carts and stacked all these boxes up on it and wheeled it across various intersections in the city to get to the computer centre. Well, wouldn’t you know it, the darn thing, we were trying to get up a curb and the thing spilled over and it took us a week to reassemble all the damn cards in the proper order. I said after that “we are going to change this thing”.

I was absolutely convinced that there had to be a better way. I got this brilliant idea, or an idea that I thought was brilliant about computer-based hardware, not just computers but what they used to refer to as ‘peripherals’, basically. You had mainframe computers and then you had these peripheral devices that were hooked up to them to do one thing or another. They were referred to as monitors and terminals because they weren’t really very capable devices, they were basically slave devices hooked to the mainframe. Anyway, I wasn’t happy with that, I wanted an offline device because – and this is the other sort of scary thought in this day of dirt cheap, hard disk, Winchester-based technology – the campus computer facility was charging us, you ready for this? Almost \$17,000 a year to rent 20 MB of disk storage on the mainframe!

JN That seems incredible!

John \$17,000 a year! But, of course, that implied that they had the liability of backing it up and being sure that it was safely stored on tape copies, typically on the big 8-track tapes.

Anyway, I said “that’s ludicrous”, so what happened? Round about the time that we finished the transcription of that first keypunch card text, I started going to these computer hardware shows. I found two different companies, they were both start-up

companies. I don't know if they even exist any longer but at the time they were these young geniuses who were working with microcomputer-based devices. They were incorporating that early microcomputer technology, those early Intel chips, the 8080, the 8-bit system and subsequently the 16-bit 8086 etc., to make these devices intelligent and programmable.

So what was the first thing that I found? I found terminals! They were intelligent terminals. Now, you have to remember that memory circuit chips and microdevices were very expensive at this time. But these intelligent terminals that I was looking at had built-in editing capability and button editing capability. The buttons were actually labelled 'insert character', 'delete character' and all that sort of thing, right on the keyboard. You could do that one 1024-character page at a time. It had enough memory to do what was more or less equivalent to one page of single spaced text, you could hold it in memory, bring it on the screen and edit everything locally.

Then I said "Well, now we need to transmit it somewhere. I don't want to transmit it directly to the mainframe computer because they're charging me by the second of online computer usage at an outrageous fee". So, lo and behold, at the very same hardware show I found a booth where they were selling a low cost magnetic tape cartridge, these were digital grade Philips cassette tapes and it had a dual tape deck. Philips, one of the few we still have around, right? It wasn't cheap, it was five grand for that particular device. But it would be cost effective for us over the long haul to enter our data and edit it offline of the mainframe UNIVAC computer.

Then, of course, I had to bring pressure to bear on the moguls who ran the campus computer system to develop spooling software that would enable us to spool the data off of these cassette tapes, in a batch mode, into the campus mainframe for processing. Now this is all before I was able to get us off the mainframe computer anyway, but in any case, I'm trying to be as chronologically correct as possible. What I managed to do was to convince these two companies (since they were small and flexible) to interface and to write the microcode which would tell the other company's tape drive to open up and receive the data that was being transmitted. That is a page of edited data and we'd just concatenate it, ok? And they were willing to do it! I think back and think that today these people would tell me to get out of their sight. You know, you ask them to do something that's going to sell five units or whatever but they were anxious enough to be willing to do it and they did.

Now, why did I do that? Well, there was an alternative device, it was called the IBM Magnetic Tape Selectric Typewriter (MTST) as it was dubbed by IBM, which interfaced its own tape drive. There was nothing else like it, it was IBM and it was not compatible with anything, obviously, nor would ever be compatible with anything. It was built into this box and connected to an IBM Selectric Typewriter with the little bouncing ball on it. They were leasing that device for about \$15,000 a year. I needed five of them as data entry stations and I had the first pilot grant from the NEH at this point and so I was able to hire staff.

Now you have to understand that my data entry problem was nothing like most of the other Humanists who were getting started in the field. Many of them actually would send their books to Hong Kong and there were these data entry services that

would simply sit there and type. All these people being paid peanuts to bash away the stuff they were seeing on these printed texts. Well, we couldn't do that because you had to be a trained palaeographer to be able to read our input and work directly with original thirteenth century manuscripts (and later ones after we completed the Alfonsine Corpus).

So I was fortunate, quite frankly. I finished my dissertation and graduated and they decided to keep me on as an assistant Professor. Someone jokingly said that the Dean kept me on because he knew that I had a quarter of a million dollar federal grant and they wanted to get their 45 cents on the dollar. That wasn't true, I actually asked the Dean myself personally and he laughed and said "no that wasn't true". They kept me on because they thought I was worthwhile; it was quite straightforward.

JN The first NEH grant was awarded around 1972?

John I can't remember the exact date, I remember drafting the proposal in '71. It might have been granted in '72, it's a lengthy process of passing through.

JN And what was your PhD thesis about?

John While working on my PhD thesis I did use much of the computer technology that we had developed up to that point. Now remember that we were doing data entry using intelligent editing CRT terminals with their own internal editing capability and interfaced to a standalone system, those dual tape drive affairs that I referred to as storage media. That duo, made by two different companies, replaced what would have been this outrageous rental from IBM for their Magnetic Tape Selectric Typewriter. Moreover, we didn't have to print out the lines when we wanted to edit, do you follow? We could do all of our editing right on this very fast CRT and when the page was edited we pushed a button and it was transferred over to tape two (the unedited version was on tape one on this dual tape drive). So, we could transfer the newly edited version to drive B and then we ended up with an edited version of the data. In any case, this was available to us when I published a version of my dissertation years later. Can you see? [He holds a book up to the webcam].

JN Yes I can: *El Libro de Marco Polo* (see Nitti 1980).

John Yes, my dissertation wasn't even on a Castilian Spanish text, it was the editing of the only extant medieval manuscript translation of the Book of Marco Polo in an Ibero-Romance tongue (Nitti 1972). It was translated into fourteenth century Aragonese and fortunately the manuscript itself still exists at the Escorial Library in Spain. I got a grant to go there and work with the original manuscript. It's a beautiful thing, it's huge, about 3 ft tall, open it's about 2 and a half ft wide and the letters are a good ½ to ¾ in. high. It was all done by hand, of course, on parchment with fancy illuminations and miniatures and what not. Just beautiful – that really got me going. I was at the Escorial for a month working with that. I had done the rough

transcription here in Madison from a black and white microfilm copy and then I took it there and actually had to make changes to the rough transcription. I can tell you right now, very often what happens is there will be secret little notes or changes that are written in the folds of the parchment and concealed by the binding. If you open it up and photograph it you have to spread the sheets apart to be able to see the notes. I found many of those, where they were actually making corrections to the text in the margins.

In any case, how do we get the edition to print in a beautiful, professionally-done typeface? Well, I'll tell you what we did. The only true typesetting was, in fact, commercially available because then typesetting devices were quite expensive. So I struck up a deal with a company based in Milwaukee, Wisconsin called Color Corp and they basically had a contract with one of those big chain stores to do all of their printing of ads and that kind of stuff. And so I said "when your machine is fallow wouldn't you like to make some money?" And they said "Sure".

So, the typesetting was very cheap, we were able to get professionally-done typesetting for probably less than a third of the cost that it would have been if we'd gone specifically to a typesetting service. These guys were doing it during fallow time. So, this book, my edition of Marco Polo was done using professional typesetting. Now, I prepared all the data input from this end and they gave me a copy of their typesetting language, that is, their mark-up language and I put the codes into the magnetic file myself, bold face, italic and whatever, point size changes and everything. It's difficult for me to talk about this sequentially Julianne because all this stuff was coming in tangentially. The technology I'm talking about was coming in tangentially and we were looking around and grabbing at whatever was available and affordable and that we figured could make it easier for us.

Now, when you look back you might say "ha ha, they did that? Who cares?" I remember that CHum asked me to write a piece and at the time I had just discovered, at another one of these hardware shows, a new device that had dual 8-inch floppy disks. Imagine that, each one of those disks would hold a quarter of a million bytes, this is nothing, right? But dual because you could edit, then go from one to the other and that device actually had a built in version of the BASIC programming language. I looked back at this article that I had written years ago and I thought "what?" I wrote this "it even has built in BASIC programming in this device and firmware!" And I even put a big exclamation mark at the end!² I had to laugh afterwards, 15 years later, I'm thinking "who cares about that?" It's that sort of a thing! When you look back on this early technology, even some of the then more sophisticated stuff, it all becomes a historical curiosity, which I assume is what you are really all about here in this thing!

²The article stated: 'One such floppy can be purchased with options such as extended core memory, or the BASIC programming language, in addition to a million characters of disk storage!' (Nitti 1978, p.46).

JN Exactly and what also fascinates me is the process of how people encountered the computing of that time and thought “I can apply this to my research”, especially considering how few Humanists used computing then.

John Well, there were many such eureka moments for me and most of it came from going to these computer hardware shows. In May of 1981 I took a show on the road for publicity purposes and was asked to give lectures about our work and the technology we used. I must have gone to 20 different American Universities, I was even invited by the University of Montreal in Canada. I went there, and I had a sort of a roadshow; my friends jokingly referred to it as ‘Nitti’s dog and pony show’. I would actually lug some of this computer equipment that we had put together, which was innovative at the time, and take it there and show how it worked. I haven’t told you yet about the other little piece of technology that I was able to work into this and that was Optical Character Recognition (OCR) at a time when the Kurzweil Scanner was very new. It cost more than \$80,000 and we couldn’t afford it.

In fact, I discovered at that time there was only one Kurzweil true OCR device in the entire state of Wisconsin. It happened to be here in the city of Madison and it was owned by a wealthy attorney who had set about to scan retrospectively all of the law statutes for the state of Wisconsin from the printed books. So I worked up a deal with him that I could use his Kurzweil and get my staff to go in during the graveyard shift when he wasn’t using it. So I hired graduate students to go there late at night and we trained them on the Kurzweil device.

I was already thinking about how we were going to create a body of word definitions in Spanish for the purposes of the *Dictionary of the Old Spanish Language*, right? It’s a different issue. We were simultaneously developing software to bring together all of the lexicon that we were compiling in the *Dictionary of the Old Spanish Language*. The first phase was to be the *Dictionary of Alfonsine Prose*, that’s the thirteenth century corpus, and so what do we do? Well, I figured, if we scan and get into machine readable form what was in the public domain, which was the then last edition of the Royal Spanish Academy’s *Dictionary of the Spanish Language*, a monolingual dictionary, then we could modify it to our liking and it would be our definitional canon, in effect. While it’s a contemporary dictionary of Spanish the first editions of it were created in the late eighteenth century, you know, the Century of Lights, and so it contained retrospectively huge quantities of the word forms that we were finding in these medieval texts. And so I said “Ok, let’s do that”. Well, we did manage to get the whole thing scanned and one of the big shots in the Real Academia Española was a buddy of mine, a senior Professor. In fact, I had brought one of his sons, who was and is still a scholar of Medieval Spanish, here with NEH money that I had to help me work on the *Dictionary of Old Spanish*. We were working on it, and his father liked that obviously.

When we finished scanning the 1992 edition of the Royal Spanish Academy’s dictionary I sent his father, who was one of the top two people in the Academy at the time, a copy of the machine-readable text of their dictionary. Obviously a political stunt but anyway it worked so they never frowned upon our using that text and

creating a definitional database in effect out of their dictionary. Obviously, it didn't look anything like their dictionary when we were done with it. It existed only in machine-readable form because all we needed was to be able to develop a software that would go in and grab the appropriate definitions out of the dictionary and pull them into our growing *Dictionary of Alfonsine Prose*.

JN So did you have some formal training in computer programming?

John Two programmes existed on the campus mainframe and both of them were uppercase-only type things. One was the concordancing programme I mentioned and the other was a bibliographic management programme, as they called it. This basically enabled you to create bibliographic records and sort them and index them and that kind of thing, which was handy but once again in uppercase. In fact, I used it because in addition to creating the *Dictionary of Old Spanish* ourselves, I had to create and establish a canon of the known Old Spanish manuscripts and early printed texts.

Old Spanish is considered up until the year 1501 or the beginning of the sixteenth century. So, obviously, early printed books had already begun thanks to Gutenberg in the last half of the fifteenth century and a number of printed texts were included in the corpus of Old Spanish texts. I created what we called the Bibliography of Old Spanish Texts (BOOST). So we started there and, of course, it was printed out using the yucky chain printers, all uppercase, on the campus mainframe. But the thing started to grow and it assumed a life of its own until finally I turned the whole bibliographic arm of the thing over to a famous Professor at the University of California-Berkeley, he recently retired, his name's Charles Faulhaber. One of his interests was bibliographies, so I turned it over to him, and he's turned it into a completely different thing. His much expanded work is called PhiloBiblon,³ and I guess it's still available online.

You know the other thing that I didn't really emphasise and it must not be lost sight of is that the DOSL project required first-hand knowledge and training of Medieval Spanish palaeography. Fortunately, I was teaching at the time, and continued to teach right up until my retirement, courses in Old Spanish palaeography and a surprising number of graduate students would enrol in those courses, especially given the sort of esoteric nature of them. It was from those classes that I was able to recruit many of my workers, my student help. We would have to adjust and fiddle the schedules and the like but many of them wanted to pick up some extra money in the summer months because they were Teaching Assistants, let's say, in Spanish at the University of Wisconsin during the school year but they didn't have any income in the summer. So the NEH grants provided me an opportunity to pay them a salary for the summer months. They knew that I could only hire them once they had taken

³PhiloBiblon is 'a free internet-based bio-bibliographical database of texts written in the various Romance vernaculars of the Iberian Peninsula during the Middle Ages and the early Renaissance'. See <http://bancroft.berkeley.edu/philobiblon/>

my course and therefore knew how to transcribe Old Spanish texts from the photographic reproductions of the originals.

From the very outset of the project I thought “is this going to be too much? How can I manage to input 11 million words of text from the original manuscripts unless I have a small army of people who are trained to do it?” But I came up with another solution which also involves technology and it has to do with the OCR. The true OCR is the only way we could scan the Royal Academy’s Dictionary because of the complexity of the typography.

But when we were doing the transcription directly from a photographic reproduction of a medieval manuscript, I thought to myself “now there are a number of professors out there and I know most of who would be likely to participate”. Assistant professors largely were the ones who were hungry and wanted something to do but some senior people also got involved in it. I found, in another one of these computer equipment shows, a standalone textual scanning device, except it wasn’t true OCR. Since it wasn’t true OCR it wasn’t bothering to read the letters and therefore it wasn’t going to cost \$100,000, right? It was a device made by a start-up company in Miami, Florida. I became good friends with the sales rep at the show, I mean, literally, we became good friends and he managed to convince his bosses to sell us one of these devices (which list priced for \$20,000) at cost price, about \$10,000.

So we bought one of these devices on the promise that I had to take it on my ‘dog and pony show’ and show it off as I went to these various campuses. This device used, are you ready? IBM Selectric Typewriter technology with the little golf ball, except they manufactured special little golf balls that had, under each of the letters, a miniaturised barcode. So, when you typed the text it would come out in alphabetic words and beneath each of the letters was a miniaturised barcode. That enabled this device, which had an automatic page feeder and everything, to scan the sheets. You could put 100 sheets in the hopper and it would scan them and create the digital images of the characters, the underlying ASCII code for the letters. I managed to convince the company to programme the device to transmit to those Philips cassette tape recorders that we had interfaced to what we were using at the time to do the data entry and editing. So the company that produced this scanning device, I call it that because it wasn’t true OCR, it was reading the barcodes and outputting text which I was receiving from my colleagues at various universities who were trained palaeographers in their own right. They would sit at home and I would provide them with the typewriter element, the ball, right? They would get their Deans to buy them an IBM Selectric Typewriter if they didn’t happen to have one already, you know, which is low-tech really and then they’d prepare transcriptions.

And what could I give them in return for this? The promise that our publishing house, the Hispanic Seminary of Medieval Studies, would publish, at least in microfiche form (because we were now creating microfiche as a publication medium which enabled us to publish hundreds of thousands of pages of information). We are still selling those microfiche packets where we’ve got the medieval text transcribed and its concordance printed out on microfiche. We sell them for \$10 apiece. In some

cases, 10,000 pages of information for 10 bucks. Because it was so cheap for us to produce microfiche we could afford to sell those packets for 10 bucks and we were making probably, I don't know, 75 % profit or something to feed back into the thing.

JN This is almost a prototypical scholarly crowdsourcing approach. Maybe crowd isn't the right word, more learned community...

John Exactly. We didn't have personal contact with these people. There would be long intervals of months, in some cases years, and then suddenly a stack of these specially typewritten, barcode-type texts would appear on my doorstep for scanning. So, I'd go out there myself, feed them into the hopper and dump them onto these Philips tapes and then I'd print it out. By that time we had our own high speed upper and lower case printing device, which cost us about \$5000. It was a stand-alone device which I interfaced to the tape drive. I'd play out these tapes on this printer and mail back a printed copy of what got scanned to the individual who had submitted the typed pages. Then it was their responsibility to go through and mark-up those pages for errors of theirs and for any possible scanning errors. And then they'd send them back to us with red correction marks on them and I had my grad student staff sit down and interactively, using the devices I told you about (the intelligent CRT terminal interfaced with the tape drives) make the corrections for the transcript.

I never finished the question you asked me about computer programming. I took a course in BASIC and I said to myself "this is silly, I'm going to try to get money from the government or wherever I can get it from." In fact, Mr Kasten put up some seed money out of his own pocket to hire our first computer programmer. By the way, all of my salaried programmers for the entire duration of this project, nearly 20 years, were women. And I'll tell you why. I discovered that young male computer programmers did not possess one quality, many of them were brilliant and excellent programmers, but that quality was constancy. I knew that with these gals, and I hired females from age 22 to 56, that they were there today and I could count on them being there tomorrow.

I haven't even talked to you about the programming we did to create a dictionary from what we'd come to refer to as citation slips. We modelled this thing, in broad terms, after the *Oxford English Dictionary*, so that our dictionary has, for instance, dated citations (bits of snippets of text out of a manuscript with the date of the manuscript associated with the snippet). In that sense it's not only a period medieval dictionary but it's historical, within that period. And that was all done through the programming that we developed ourselves, our own ad hoc software.

JN You developed a whole intellectual, technical and social infrastructure that supported the project. Did the OED's use of 'crowdsourcing' also inspire you?

John I still think OED is the best dictionary there is! I had been impressed with it from the very outset and I remember having read and seen stories about how the conception of OED came about and the idea that they actually had a bunch of non-

technical people sitting at home and writing down what they found. Their job was to read text and pull out words that they thought were neat and hadn't yet been documented, or whatever, and then they had to write them on snippets of paper. That intrigued me. I said "well, let's see, how we can do that and go one step further. We can capture their keystrokes instead of having them send us a bunch of snippets of paper, right? We can actually have them send us machine-readable pieces of paper." That's what inspired me, the analogue in a technologically more sophisticated and facilitating manner.

JN Can you please reflect a little more on the whole process of how you conceptualised, designed and implemented this whole infrastructure?

John Well, it started out sort of helter skelter. I was learning as I went along, basically. This is going to sound terribly immodest, I don't mean it to be, it just happens to be true. I didn't at any point say "I can see that computers can create a concordance from this albeit primitive looking machine." I didn't have that capability. No, I said "is there computer technology available, both hardware and software, that can do what I really want to do ... my dream world, what's my dream world?" I was driven, and I think that's the proper word, I was driven by this notion that there had to be a better way. So, every time in the process of integrating all these things, people and equipment, I felt as though it could be done a better way, whatever particular aspect it was I was dealing with. I would set about to try to see if there was a different, better and improved way to do it. Of course, I was fortunate that during the course of this 20-odd year odyssey, technology itself was not static, obviously, so there were new devices coming on the market, there were even new software packages.

I haven't told you yet about the software that we wrote. I should have gone back because mixed in with all this stuff was John Nitti assembling PCs before there were PCs. There were kit computers coming out in California, these garage built kit computers and they would send you the components. The first one I built was driven by a little 8-bit, Intel 8080 chip and it started with 64 K of RAM, which cost a lot of money. It came in what looked like a mahogany window box, something you'd plant flowers in. It was about a yard long and quite narrow because at the end of it you had two of these 8-in. floppy disk drives. Those were the days when the floppy disk actually did flop. I built the damn kit computer and I actually started to try to offload some of the sorting procedures that we did in organising words. Believe it or not, and it was a pain in the tail because the storage capacity was limited to these two quarter of a million byte floppy disks. One floppy disk you'd stick in there with text and you'd sort the stuff you wanted to sort and it would write the output to the other. We developed our own little sort algorithm to run on that early microcomputer. This is before IBM PCs and TRS-80s and that sort of thing. That was my favourite old thing, I should have kept that.

Then I connected with another outfit in California. You're too young to remember the battle between the two big microcomputer operating systems, the CP/M

(‘Control Program for Microcomputers’; see Kildall 1982) and MS-DOS. Why, Gates was very lucky because when IBM was planning to release its 16-bit PC and were looking for an operating system they chose Bill Gates’ MS-DOS. As a result, everything was modelled after that operating system and then Gates and the Microsoft Corporation, of course, started producing subsequent iterations of that plus Windows. In any case, in the early days you bought the components and you had to build the computer yourself. Because there initially was no multi-user operating system, I later migrated to MP/M, the true multi-user, multi-tasking version of CP/M. In fact, the MS-DOS notion in IBM was basically about networks, so you would network a series of PCs attached to some central PC, right? Well, during all of this, I build an eight-user multi-user system, a true multi-user system where there were eight terminals. We used it for years. These were just dumb terminals interfaced to the computer I had built that was running the multi-user version of CP/M, called MP/M. Now people say “What? What’s that?” Most people don’t even know what that is because it didn’t happen to be chosen by IBM as their favourite operating system. So, in any case, we used that system as a subsequent data editing station. I had 8 students simultaneously entering data to a kind of a hard disk device, early Winchester technology. I concatenated four, are you ready for this? Four 80 MB (that was a lot then) disks drives together in one enclosure and interfaced it to this multi-user computer. So we actually had a sizeable chunk of hard disk. Now remember, I was no longer paying the \$17,000 a year to the campus mainframe people for 20 MB of hard disk.

JN You just mentioned your vision, as it were, of your perfect world. It just occurred to me that I didn’t actually ask you to describe that.

John My perfect world was also dynamic because I went about it in the following way. I said “OK, there are particular tasks that this project of ours needs to be able to do, using computing technology and computer-related technology”. They included, of course, computer-based typesetting. I went with that and had the sub-contract early on with this company I told you about. But I wasn’t content; I wanted to be able to do our own typesetting, in house. We finally had some sophisticated output printing devices coming on the market. Now, of course, I did buy a Lexmark all-in-one multi-purpose printer for \$50, which is as much as the inkjet cartridges for it cost. Of course, it produced good typographical quality stuff, not necessarily the finest typographical but certainly suitable for reproducing and publishing in books.

So we then went to “how are we going to be able to do the typography ourselves on a microcomputer?” Well, it turned out that Donald Knuth had invented and wrote for Unix mainframes the TeX typesetting language (Knuth 1979). He turned it over to public domain and as soon as he did little companies started producing, in this case early on, MS-DOS-based versions of TeX for 200 bucks. I bought the complete typesetting capability, with more sophistication than I ever, ever, ever imagined we could use, because it could also typeset sophisticated mathematical formulae and that sort of thing (in fact Knuth had designed it for that purpose originally). So then

I said “now we’ve got the core software” and we actually bought the source code of that package. And my programmers, the different ladies I was referring to earlier, could in fact develop and interact with that typesetting language in such a way as we could create our own output in electronic form: typeset pages including the changing of the running heads with the alternating pagination. We were controlling the typesetting software itself. We were sticking our noses into that typesetting software and saying “this is what we want you to do, dammit!” And it did!

In fact, I typeset the *Dictionary of Castilian Prose of King Alfonso X* right here at my home in Madison in 2002. I printed the entire dictionary on a low-cost, high quality laser printer which cost under \$1000. I had already transferred the Hispanic Seminary (the publishing house) to John O’Neill in New York, right? I sent him the camera ready copy and he negotiated a contract with a printing outfit and that’s how we were able to sell this.

Well, I think as a result of all this, whether correctly or not, I guess I got the reputation of doing what I did best and knowing what I was doing. So Helen Agüera at the NEH (see Chap. 10) and I used to go on site visits a lot. We’d meet and say “we’re going to Yale this time” and we’d convene there for a site visit of a humanistic project that wanted to employ computer technology, especially research tools projects, which was Helen’s area. We had great fun! Stuff Ivy League professors weren’t necessarily happy to get my advice, but they got it anyway!

JN So this brings us back to the question about scholars who were not using computers in their research and the sense you might have of their views on Humanities Computing?

John I’m glad you made the link, it’s a good one. Basically, I have to start out by underscoring the fact that some 3 months before his death at age 94, my mentor, Mr Kasten, was still working with me on a daily basis on completing the *Dictionary of Alfonsine Prose*. We were sitting side by side, with me at the computer terminal and Mr Kasten working through the proofed copy of the dictionary pages, effecting the corrections he had found. I would make the electronic change and go on to the next page. Mr Kasten was an incredible fellow, he didn’t know anything about computer technology, nothing! Zero! He knew less than I knew at the beginning. But I was able to start the computer experimentation, thanks to Mr Kasten because he was bankrolling me, particularly before we got anything from the NEH pot. In total, the largest funding we got was from NEH but we also got \$300,000 worth of matching money from the Spanish Government. I was in fact named as a visiting Fulbright scholar to the University of Salamanca, which is important for this purpose, not because of me but because I went there after we had developed all the microcomputer-based lexicographic software.

I installed all the software gratis on the computers of my colleagues in Medieval Spanish at the University of Salamanca, which is one of the oldest universities in Europe. I was there for 3 months to teach them how to use our software, which I installed on three PCs they purchased for the purpose. They were three ladies, two professors and the wife of a Spanish professor of English, and I jokingly referred to

them as *mis tres Marias*, ‘my three Marys’, because Spanish Catholic ladies names frequently start with “Mary”. We had a heck of a good time there and I taught them all how to use the stuff and they, in their own right, created two separate dictionaries, big monster dictionaries. The more important of the two was a *Dictionary of Medieval Spanish Medical Texts*. So the people who didn’t use computers, right? These three ladies had never looked a computer in the eye. I mean their campus was bringing in PCs for the offices and what not so they probably were writing letters or something. But they got into this with both feet. As a matter of fact, in terms of the chronology of the publications, their medical dictionary actually came out before ours did. I did all the typesetting and everything here, again, in my house. I typeset that dictionary and they paid for a courier to hand deliver the typeset pages to them in Salamanca.

So I had two long-term experiences with older people, in this case, scholars, researchers and professors. You might expect some resistance, you know. I guess that’s why you raise that question and the answer is yes and no. The people here at the University of Wisconsin, the older people were delighted with this stuff. As soon as I was using PC-based technology Mr Kasten said “buy me one of those things, I want to take it to my house”. He was, by the way, a fantastic typist (I’m talking about conventional, manual typewriters), even as an old man. So I got one for him and sat him down with it, taught him how to use it, the word processor and all that stuff. He was actually typing in his bibliography, his own library collection actually, typing and entering. So it was a delight to see him with absolutely no hesitation or no griping about this technology stuff.

Another one I think you’ll get a kick out of, which is very important, was a delightful and brilliant scholar by the name of Frederic Cassidy. Professor Cassidy was Mr Kasten’s contemporary and also died in his 90s. He was the founder and the editor of the *Dictionary of Regional American English*. I was talking to him one day and he had invited me to serve on that dictionary’s Board of Advisors (it wasn’t that we met all the time, but he happened to stop by to see Professor Kasten). When I showed him what we were doing with the computers in terms of data entry in particular (you remember I mentioned to you those intelligent terminals and the tape drives that we had interfaced at that point?) He said “I want those”. So, he incorporated that same hardware technology and even hired away one of my female computer programmers to work for him programming the *Dictionary of American Regional English*. She was on top of everything we had been doing. His dictionary was considerably different in its nature and scope obviously, so there wasn’t any software transfer. But the hardware, of course, needed to be able to handle these great gobs of data and somehow capture the keystrokes instead of being fleeced by the campus computing facility for the online services.

I then started to give lectures on our computer-based techniques at various universities. I would meet people who would come up afterwards to talk and there were some instances of people who were Humanists. Though I don’t think it was simply an age proposition, quite frankly, the youngest ones were on board from the get go. They saw it was not only the future, it was the present, so they wanted to know. The International Congress on Medieval Studies held annually at Kalamazoo, Michigan

is the biggy of international Medieval Studies. I don't know how, but Professor Otto Gründler, who ran it for some 34 years, until his death in 2004, had gotten a hold of my number or something and called me and said "would you like to chair an ongoing session on computers in the Humanities here at the Kalamazoo conference every year?" I said "I suppose so". I signed up to that and I did it for some 5 years, and I was told, though I don't know if this is true or not, that it was the most heavily attended session at the conference, which seemed a bit much to me.⁴ But you can see it spanned all the fields so somebody who wasn't interested in Chaucer or Alfonso the Wise but was interested in the application of technology to their own research would attend.

It went on until the fifth year when we had PCs hitting the market, the Apples and the TRS-80s and that sort of thing. I sort of got this sinking feeling that many of the people were attending not because they were interested in learning anything new, but because they wanted me to confirm that they had made the right choice in buying their PC. You know, it got to be brand specific, shall we say. I figured this session had outlived its usefulness now so I stepped down.

By the way, I found this little card right here [holds card up to web cam]. John Nitti, Professor of University of Wisconsin is on it, and it is an admissions card to the 1981 National Computer Conference, which was the big conference for computer types, not for Humanists. This was 1981, it was being held in Chicago and I was invited to present what I was told was the first talk offered by a Humanist at the National Computer Conference. I can't swear to that, that's what I was told by the people who invited me to do it. I gave a presentation and what excited these people, I think, was the fact that I was incorporating all sorts of computational and related hardware that demonstrated how this technology, in particular the hardware, could be applied to humanistic research. My session was well attended. I was surprised, since I figured there'd be four or five people sitting there, you know, out of thousands of people but I was in disbelief! You know, who cares about this? Particularly among computer types, you see.

JN You already anticipated the issue of resistance. So did you encounter any significant resistance?

John I'm a very pragmatic person and stubborn as hell. I went into the thing saying "I know what the devil I want to do and so the only thing that's going to impede me from doing this is money. The moment the money valve is shut off I can't do anything more". I'm not independently wealthy so I couldn't do it myself. Fortunately, I had that angel in Mr Kasten, who was willing to put up tens of thousands of dollars. He was an elderly man already and he was at times as giddy as a kid when I introduced him to a new gizmo.

⁴The International Congress of Medieval Studies 'Archive of Congress programmes' lists Nitti as having participated in three such sessions at the 16th, 17th and 18th Congresses, from 1981 to 1983. See http://scholarworks.wmich.edu/medieval_cong_archive/

There is stuff I haven't even told you about. Another gizmo I interfaced into all this was a product that was initially made to order for the United States Patent Office. It was a computer-driven microfiche retrieval and display play unit that had a carousel in there. This was the time when we were doing microfiche output. You could load it up with microfiche and each frame had its own address and you could build an address table for everything that was in there.

We would bring up, in one second, a photographic colour reproduction of a manuscript page (that corresponded to a page of transcription that we had just done) to do final checking against the original manuscript. "You see", I'd say to Mr Kasten "I want this machine". He'd say "well, what does it do?" and I'd explain it to him and he'd say "ok, how much do you need." I'd say "\$10,000" and he'd say "ok" and write me a cheque. This was very important because with NEH funding you have to have pre-budgeted everything. So if there was a new gizmo and I liked it and wanted it because there was a place for it in the project I couldn't go and rob the NEH grant that I had because everything had already been allocated. So I'd go to Mr Kasten and he'd provide me with the money to do it. I had great fun doing it, as a matter of fact, as you can well imagine.

JN You've already mentioned a couple of things relevant to this question but I wanted to ask about your encounters with the Humanities Computing community or conference scene

John It is funny because when the whole thing started, I was, I guess, one of a handful of pioneers. I hate to use that word as it sounds self-serving but other people use that in my connection. In fact, instead of having existing bodies and groups that met on a consistent basis that I could go to, I was throwing parties, as it were, for Principal Investigators in projects that were purporting to use or were using computer technology. They would come to Madison, to the Seminary of Medieval Studies where I hung out, and where we basically owned three quarters of the 11th floor of our tower building. We paid for every square foot, by the way, as the Dean kept reminding me in the indirect cost that they would rake off the top of the NEH fund.

I think we had at least three such events and they were not terribly formal. I just contacted them and said "why don't we have a brainstorming session? Come to Madison and I'll arrange for your hotel room". It was always six, at the most, seven Principal Investigators from various projects around the country; they came from as far away as California. When I finally got everything established I knew exactly where I was going with the project. This, again, is going to sound terribly self-serving on my part ... let's put it this way, my need to communicate to other people was being ratified by various universities where I talked about the project. Then the fun part was not me giving the talk, the fun part was afterwards. I'd hang around and people would come up in great droves with all sorts of interesting questions. I don't know if I answered them all successfully but it was fun.

In fact, there was a sort of a clique-ish group of us, perhaps that's not the right word, there was a group of Principal Investigators, myself included, in the early

days of the NEH's first willingness to offer funds for incorporating computer technology into Humanistic research. That group of guys, it was all guys at that point, were all asked by the Endowment to draft the first guidelines for Principal Investigators in the Humanities seeking to incorporate computer technology into their research. It's been redone countless times, I'm sure, since then.

JN Many thanks indeed for your time and this fascinating interview

References

- Jover, F. G. (Ed.). (2002). *Two generations: A tribute to Lloyd A. Kasten (1905–1999)*. New York: Hispanic Seminary of Medieval Studies.
- Kasten, L. A., & Nitti, J. J. (Eds.). (2002). *Diccionario de la prosa castellana del Rey Alfonso X*. New York: Hispanic Seminary of Medieval Studies.
- Kildall, G. (1982). CP/M: A family of 8- and 16-bit computer operating systems. *Education*, 102(3), 211–219.
- Knuth, D. E. (1979). *TEX and METAFONT: New directions in typesetting*. Bedford: Digital Pr. and American Mathematical Society.
- Nitti, J. J. (1972). An edition, study and vocabulary of the unique Aragonese book of Marco Polo translated by Juan Fernández de Heredia. PhD thesis. University of Wisconsin-Madison.
- Nitti, J. J. (1978). Computers and the old Spanish dictionary. *Computers and the Humanities* 12(1–2): 43–52: 46.
- Nitti, J. J. (1980). *Libro de Marco Polo. Aragonese versión*. Madison: Hispanic Seminary of Medieval Studies.
- Nitti, J. J., & Kasten, L. L. (Eds.). (1997). *The electronic texts and concordances of medieval Navarro-Aragonese manuscripts*. Madison: Hispanic Seminary of Medieval Studies.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

