

Entrevista: PLN em Português – As origens

Uma conversa com Diana Santos (Linguateca) e Maria das Graças Volpe Nunes (NILC)

Aqui você lê uma entrevista com duas pioneiras do PLN/linguística computacional de língua portuguesa, mas cujo papel vai além do pioneirismo. Graças ao esforço de ambas, cada uma de um lado do Atlântico, foram criadas as condições para que o PLN em português pudesse avançar.

Propus as mesmas perguntas, e as respostas de Diana (que incluem um glossário criado por ela para compreendermos melhor o contexto português) vêm por último.

Conversa com Maria das Graças Volpe Nunes (MGVN)

1. Como foi seu primeiro contato com PLN? Como o PLN chegou a você?

MGVN: Foi mais ou menos em 1985 e foi minha porta de entrada à área de Inteligência Artificial (IA). Já era professora de computação no ICMC-USP, São Carlos, quando o Prof. Antônio Eduardo Costa Pereira, hoje na Universidade Federal de Uberlândia, MG, nos apresentou a linguagem de programação PROLOG, que era a “concorrente europeia” da linguagem LISP, amplamente usada na academia americana na ainda incipiente área de IA. Nessa época, os sistemas especialistas eram o principal exemplo de IA: aplicações dedicadas a um domínio bem delimitado (por ex., uma especialidade médica), com conhecimento representado num dos paradigmas da área, e desenhadas para interagir com humanos. Nessa interação, ocorria o PLN da época: interpretação de perguntas muito simples, quase codificadas, ao sistema. Como resposta, via de regra, apenas o necessário para contemplar a pergunta. Nada que se assemelhasse ao que atualmente os sistemas interativos podem fazer.

As características de uma linguagem lógica como Prolog, pela qual todo conhecimento é representado na forma de fatos e regras, e todo comando equivale a um teorema a ser provado, tornavam relativamente simples o processamento da língua para aquele tipo de aplicação. E foi assim, projetando interfaces de sistemas inteligentes com humano, que o PLN chegou até mim.

2. Como é a sua relação com a língua portuguesa? Afinal você é de engenharia, e temos o folclore de que quem é de exatas não gosta de humanas...

MGVN: Fazer um curso de exatas, no meu caso, não foi uma decisão muito natural. Lembro-me de ter feito um teste vocacional, pois nunca me foi muito claro o que gostaria de estudar. Talvez porque sempre gostei de todas as disciplinas e, para a minha geração, o importante era escolher “o que se queria ser” como profissional. Hoje, muitos escolhem a carreira pelo gosto que têm por uma ou outra disciplina, e não pelo o que terá que fazer como profissional.

Sempre gostei bastante de ler e, hoje em dia, também me dedico à escrita. Sou daquelas pessoas que admira quem fala muito bem e que torce o nariz para os que cometem muitos erros ao falar ou escrever. Mas confesso que foi depois de começar a atuar em PLN que passei a me relacionar com a língua portuguesa de maneira especial. Descobri que não a conhecia como imaginava. Descobri seus encantos e também suas armadilhas. Tentei domá-la muitas vezes, mas hoje estou mais para fã do que para algoz. Meu passatempo atual é descobrir os limites que ela impõe ao seu processamento automático completo.

3. Tanto o NILC quanto a Linguatca são ou foram criados\liderados por pessoas (mulheres) de computação/engenharia. Como foi a aproximação com os/as linguistas?

MGVN: Aqui no Brasil, a iniciativa de um centro de PLN em instituições de exatas não surpreende, já que o PLN surgiu mesmo no âmbito da computação. Já a liderança do NILC, na época de sua criação, era dividida entre mim e o Prof. Osvaldo Novais de Oliveira Jr., do Instituto de Física da USP, São Carlos. O grupo foi criado como consequência de um grande projeto — um revisor gramatical inédito para o português — portanto, já sentíamos a necessidade de especialistas na equipe. Olhando em retrospectiva, considero que acertamos ao convidá-los como parceiros, e não como profissionais contratados. O grupo é chamado de núcleo exatamente por congrega diferentes grupos de pesquisa em linguística e computação. E os estudantes bolsistas que desenvolviam os projetos eram, em sua maioria, oriundos dos cursos de linguística ou letras. Temos orgulho de que muitos deles se tornaram profissionais nessa área, aqui e no exterior, bem como pesquisadores em PLN, e puderam introduzir essa área em vários cursos de linguística.

4. Algo chegou a surpreender você (de forma positiva ou negativa) no diálogo com os/as linguistas, ao longo dos anos de trabalho “conjunto”?

MGVN: Muitas surpresas e sempre positivas. Só presenciei muito respeito mútuo e as dificuldades sempre foram sanadas naturalmente.

Um problema recorrente que ambos os lados sentem é quanto à rigidez dos sistemas universitário e de pesquisa brasileiros no que diz respeito à interdisciplinaridade entre exatas e humanas. Enquanto o tempo tem mostrado as vantagens da multidisciplinaridade nos projetos de pesquisa, relaxando algumas dessas restrições, a formação no nível de pós-graduação ainda oferece muitos obstáculos. Assim, é difícil — mas não impossível — que um linguista cumpra todos os requisitos de um mestrado ou doutorado em PLN num programa de computação, por exemplo. Mas acredito que haja uma tendência de mudança, já que o PLN não é único nesse cenário; atualmente, são inúmeras as áreas que necessitam de outras para a sua evolução.

5. Como chefe de laboratório/equipe, o que você diria para jovens linguistas que desejam se aventurar na área? Para você, qual seria o perfil ideal da pessoa de Letras que integra um grupo interdisciplinar de PLN?

MGVN: Em primeiro lugar, que, mais do que uma aventura, se trata de um caminho alternativo muito interessante frente aos outros de que se dispõe. Fazer PLN não é apenas construir ou possibilitar a construção de sistemas com habilidades linguísticas como (semelhantes, iguais?) as nossas. Fazer PLN é ter um outro olhar para a língua; descobrir seus padrões e suas infinitas possibilidades, sua forma, estrutura, e também tudo o que foge dela; enxergá-la sob o ponto de vista da máquina, pensá-la formal e descobri-la fora de controle. Enfim, é ver a língua fora de seus padrões usuais.

E minha primeira observação seria: seu papel não é fazer serviços para aqueles que desenham e programam o sistema. Você está aqui para ajudar a pensar todo o problema, indicar caminhos e limites que o programador desconhece (mas que, como falante, supõe que conhece), ajudar na construção e na avaliação da solução. Não abraça a nova área como se fosse um mundo desconhecido a lhe ser apresentado por terceiros; explore esse mundo ao lado deles, iluminando caminhos que só os linguistas conhecem. Essa parceria só será frutífera se os papéis de linguistas e informatas estiverem no mesmo patamar de importância.

6. Como você vê o PLN em português hoje, comparando com os anos iniciais? E o que você esperaria daqui para a frente?

MGVN: Mudou muito. E não só para o PLN, e não só para o português.

O PLN de português, no Brasil, sempre esteve atrasado em relação ao inglês (língua mais desenvolvida na área) por vários motivos: começou bem depois; demorou a ter uma comunidade de

tamanho e expressão capazes de alavancar a área; essa mesma comunidade inicial, em sua maioria formada no exterior, sofreu muitas baixas devido às dificuldades de formar e manter grupos em suas bases; levou tempo para o processo de divulgação da área em grupos de linguistas e sofreu dificuldades inerentes à formação de grupos interdisciplinares; sofreu por escassez de recursos financeiros para projetos que acelerassem o progresso da área, entre outros fatores.

Mas a comunidade, embora pequena, organizou-se rapidamente e, tendo se juntado à comunidade de Portugal, logo criou projetos conjuntos, um evento científico comum — o PROPOR, desde 1993 — e uma parceria que se mantém até hoje. Internamente, têm sido inúmeros os projetos entre diferentes grupos de pesquisa espalhados pelo país. As parcerias deram origem a eventos científicos nacionais sólidos, como o STIL e todos os seus eventos- satélite especializados: Jornada de Descrição do Português (JDP), *Workshop* de trabalhos de IC (TILic), Festival Sintático. Desde logo surgiram, como novos parceiros, as empresas tecnológicas e de serviços, com resultados relevantes na forma de aplicações, recursos e ferramentas linguístico-computacionais para o português. Grandes projetos foram essenciais para a formação de recursos humanos treinados que têm multiplicado a comunidade Brasil afora.

Vinte anos atrás, os sistemas de IA — e de PLN — eram feitos por meio da representação simbólica: no caso do PLN, da língua e seus fenômenos. Além disso, seus algoritmos refletiam algum tipo de raciocínio tido como semelhante ao humano. Tudo isso tornava acessível o processo de construção das respostas do sistema e, portanto, possibilitando detectar e solucionar eventuais erros.

Com o avanço tecnológico do *hardware* e dos diferentes métodos estatísticos e neurais para a aprendizagem de padrões e previsões por modelos, a representação do conhecimento em quase todos os sistemas de IA, incluindo os de PLN, foi substituída por modelos aprendidos a partir de imensas quantidades de dados (exemplos) — *big data*. Com isso — e seu *Linguística computacional* demonstra muito bem — o maior trabalho dos cientistas de PLN tem sido projetar e anotar grandes *corpora*, bem como avaliar o desempenho dos sistemas construídos a partir deles.

O PLN de português — assim como o de outras línguas — tem se beneficiado da tecnologia atual por aprendizagem, pois, à medida que os algoritmos estão disponíveis, e bastam *corpora* para ensiná-los, todo o atraso pode ser minimizado se o esforço para a construção de *corpora* for concentrado. Parece simples? Não é. *Corpora* sem anotação ensinam, mas não muito. *Corpora* anotados geram melhores modelos da língua, podem prever muito melhor. Acontece que a anotação de *corpora* imensos requer ferramentas de anotação eficazes, e elas fazem tarefas simples e sofisticadas: lematização, anotação morfossintática (PoS tagger), sintática (parser), de papéis semânticos etc. Atualmente, percebe-se um esforço para melhorar as ferramentas do português — elas próprias frutos de aprendizagem automática — a fim de que o esforço maior na construção dos *corpora* para aprendizagem seja mais da seleção dos exemplos e do processo de anotação semiautomática, ou seja, com a participação humana apenas para revisão, adjudicação e avaliação.

Longe de ser um trabalho puramente técnico e volumoso, preparar os *corpora* para a aprendizagem de máquina é uma tarefa de extrema relevância para todo o sistema de IA. No caso do PLN, em particular, o impacto do modelo da língua em muitas aplicações atuais (como as que processam a linguagem das redes sociais para prever o comportamento, a personalidade etc.) é maior do que o refletido pelos índices de acerto/erro avaliados em laboratório. A língua reflete pensamentos, reflete o mundo, e modelos equivocados podem provocar danos imprevisíveis.

É difícil saber quanto tempo durará esse modelo de sistema inteligente ou se ele é definitivamente a solução para os desafios de se emular o comportamento humano nas máquinas. O fato é que essa aparente solução final traz novos e importantes problemas. Atualmente esses modelos são incapazes de se justificar, ou seja, não conseguem explicar como chegaram a determinado resultado; não há um raciocínio associado; são caixas-pretas. Por outro lado, esses sistemas são cada vez mais autônomos, tomam decisões sem recorrer a humanos. E o dilema está posto: na eventualidade de

uma consequência indesejada ou criminosa, quem se deve responsabilizar? A discussão sobre a ética dessa tecnologia inteligente que já permeia toda a nossa vida está na pauta do dia. Em particular, a comunidade do PLN tem muito a dizer no que se refere aos dados linguísticos, afinal, grande parte desses sistemas autônomos toma suas decisões a partir de alguma forma linguística.

7. Outras coisas que você queira dizer...

MGVN: (já falei demais 😊)

Conversa com Diana Santos (DS)

1. Como foi seu primeiro contato com PLN? Como o PLN chegou a você?

DS: O meu primeiro contacto com o PLN foi na universidade, nas cadeiras de inteligência artificial com o João Pavão Martins e o Ernesto Morgado, que voltaram ao Técnico (“Técnico” significa coisas diferentes no Brasil e em Portugal, e por isso está no *glossário* ao fim da entrevista) depois de se doutorarem em Buffalo, New York, e entusiasmaram quase metade dos meus colegas (de engenharia eletrotécnica e de computadores), que qual enxame à volta deles queriam fazer mestrado com eles (e eles tiveram de recusar imensos). Sobretudo o Pavão Martins, com os seus olhos verdes, o seu fumar cachimbo e a sua voz pausada, falavam de algo mágico, o futuro, e nós “caíamos todos como tordos” (Para que não fique aqui a ideia de que ele seria mais encantador de serpentes do que cientista, devo dizer que foi um dos melhores professores que jamais tive, com grande rigor e honestidade intelectual, e sempre pronto a partilhar muito do seu tempo com os alunos interessados. Organizou, por exemplo, um círculo de leitura de artigos ou livros sobre inteligência artificial, em que iam desde alunos de licenciatura a doutorandos).

É preciso desde logo dizer que a área dele não era o PLN, era a representação de conhecimento, mais especificamente a revisão de crenças (*belief revision*), mas eu imediatamente me apaixonei sobre o PLN, sobretudo a semântica. Tivemos uma cadeira de mestrado de PLN baseada num livro (não sei agora que livro era), em que cada aluno (ou grupo de alunos) dava uma aula sobre um dos temas. O meu tema ou foi a semântica ou a pragmática, ainda me lembro da excitação associada à frase “O rei de França (não) é careca”, que não era nem verdadeira nem falsa. Enfim, abriu-se um mundo para mim, que sempre me tinha interessado pela língua e pela gramática.

2. Como é a sua relação com a língua portuguesa? Afinal você é de engenharia, e temos o folclore de que quem é de exatas não gosta de humanas...

DS: O português sempre foi uma das minhas disciplinas preferidas, além da história e da matemática. Não só porque sempre adorei ler, mas porque gostava imenso de gramática: orações relativas, complementos circunstanciais etc. O português era, como alguns professores diziam, “a matemática das letras”. Até não me importei nada de dividir as orações dos *Lusíadas*, coisa que em geral as pessoas todas diziam que era assassinar a obra, e que era muito discutido se era pedagogicamente correto, blá blá...

O “folclore” que eu tenho, aliás, é o contrário: são as pessoas de Letras que não gostam de ciências. Todas as minhas amigas e amigos que foram para ciências gostavam de literatura, de filosofia, de línguas... só história é que em geral não gostavam (e nem todos, claro). Ao contrário, a maior parte

(eu até diria todas) as pessoas que iam/foram para Letras era porque tinham (sobretudo) problemas com a matemática. E também não gostavam de química nem de física.

Não querendo generalizar mais do que à minha geração (anos 1970 em Portugal), a divisão era entre os que gostavam de tudo e aqueles que só gostavam (ou só serviam) para Letras. Era antes uma divisão intelectual, potenciada pelo facto de haver vários intelectuais portugueses polivalentes (muitos escritores eram médicos, um dos poetas mais badalados, o António Gedeão, era professor de química), além de haver muito a ideia de construir um mundo novo (a seguir ao 25 de Abril) e, portanto, o que era preciso era engenheiros e agrónomos e cientistas — pessoas viradas para o progresso.

Enquanto as Letras pareciam mais bolorentas e viradas para o passado — além de terem muito menos saídas profissionais: e daí, embora a história tivesse sido sempre a minha disciplina preferida, quando chegou a hora de escolher... Um parêntesis: foi só aos 17 anos que tive de escolher, porque tive, no curso complementar do liceu, a possibilidade de escolher cadeiras que davam para ir para Engenharia ou para Letras: no meu liceu, em 1.000 alunos, éramos só duas que tínhamos física, matemática e história (cadeiras escolhidas), além das obrigatórias para todos os alunos (português, filosofia e introdução à política) — digo, quando chegou a hora de escolher, e embora rebolesse muitas noites na cama sem saber o que decidir, escolhi ir para o Técnico. As razões, além das políticas: virar-me para o futuro, ser útil, foram sobretudo as seguintes — e aqui digo que são as que eu arranjei para mim, não digo que sejam certas: não queria ser professora, o que era a saída de 99% dos que tiravam história; e a ideia de que era mais fácil estudar história “por fora”, como *hobby*, o que seria impossível se cursasse história, não poderia tirar engenharia por fora, porque era preciso laboratórios, computadores, material, sei lá mais o quê.

Mas o que é certo é que, logo que entrei para o Técnico, adorei, e embora continuasse a ler livros de história, identifiquei-me absolutamente com aquele ambiente e com aquele grupo de colegas e professores. Mais uma vez, e como estou a dar esta entrevista sobretudo para pessoas de Letras, tenho de enfatizar que, embora todos nós (ou alguns) estivéssemos muito interessados na matemática, na física, na eletricidade, tínhamos professores excepcionais com uma grande cultura geral e que, além da matéria, nos davam assuntos para pensar e interesses gerais — é a vantagem de pertencer a uma escola de elite — não é tanto pela matéria ou pelas instalações, que não seriam famosas (e a comida da cantina era terrível), mas sim pelo contacto com os melhores cientistas e engenheiros do país e pelos exemplos que eles nos davam (o que em inglês se chama *role models*). Só para dar dois exemplos, o nosso professor de propagação guiada (cabos coaxiais, antenas), Abreu Faro, já sexagenário, falava no anfiteatro da necessidade da angústia como motor de conhecimento (além de ter escrito livros excelentes sobre a matéria), e explicava (ou era sabido) como tinha conseguido construir um prédio para pesquisa no *campus*, com metade do orçamento: tinham mantido o plano original, e construído apenas os primeiros quatro andares. Mais tarde, com outros governos e outros fundos, tinham completado a obra. E o Mariano Gago, físico de partículas jovem, que vinha do CERN e que nos perguntava que autores nós gostávamos de ler — li o James Joyce por causa dele — e que dava umas aulas de mecânica geral absolutamente extraordinárias.

Por isso, quem pensa que um curso de engenharia eletrotécnica é uma chatice, não faz ideia. Claro, eu não posso comparar com os cursos de Letras da altura, mas as minhas colegas de Letras nunca me contaram nada tão interessante.

Mas de qualquer maneira, eu talvez não seja um exemplo muito normal em relação à dicotomia Letras-Ciências, porque uma das causas psicanalíticas pelas quais eu (também) não quis ir para Letras era que a minha família era toda de Letras, e eu queria afirmar-me como diferente.

3. Tanto a Linguatca quanto o NILC são ou foram criados/liderados por pessoas (mulheres) de computação/engenharia. Como foi a aproximação com os/as linguistas?

DS: Esta pergunta é um pouco difícil, porque mistura a questão do género (feminino/masculino) com a questão da formação universitária. E porque naturalmente não houve uma aproximação de disciplinas, mas sim de pessoas, pessoas muito diferentes. Para poder responder de forma honesta sobre essa aproximação, tenho que dizer que eu me aproximei a três grupos de linguistas, e com resultados e consequências muito diferentes:

— O primeiro caso foram os alunos do curso de linguística, que tinha acabado de abrir na Faculdade de Letras e que vieram aos montes para serem estagiários no primeiro projeto de PLN que eu liderei, de tradução automática. Foi muito positivo e, com alguns deles, ainda mantenho uma relação de amizade ou pelo menos contacto de maior respeito mútuo, como é o caso da Anabela Barreiro, da Regina Reis ou do Rui Marques.

— O segundo caso, que foi terrível, foi o contacto (ou falta dele) com (a maioria d)os professores da Faculdade de Letras (da Universidade de Lisboa) de Linguística, que viram esse projeto como um competidor, e usaram todas as armas possíveis para o deitar abaixo, falando mal, caluniando, e dando razão ao que se dizia na minha casa de que “a Faculdade de Letras era um ninho de víboras” — ao contrário do Técnico em que, claro, que também havia inimizades, mas não o ambiente de feudos e clãs e de intrigas... Não vou aqui entrar em pormenores, nem fica bem estar a dizer mal de pessoas que já morreram, mas a grande mentora dessa “guerra” ao INESC-IBM foi a Maria Helena Mira Mateus, professora catedrática e chefe do ILTEC, que era o membro da EUROTRA, e a quem eu me dirigi com uma ingenuidade absoluta, pensando colaborar para o bem de todos. A esse propósito lembro-me também do choque do meu chefe na IBM, o (Eng.) Ferreira Pinto, que ainda ficou mais chocado do que eu com o comportamento dela, e que disse que não imaginava que uma professora catedrática pudesse comportar-se assim perante uma jovem como eu.

— O terceiro caso foi muito mais positivo e levou a uma colaboração de vários anos, e de facto, a maior parte das coisas que aprendi sobre corpos foram-me ensinadas por essa equipa do CLUL (Centro de Linguística da Universidade de Lisboa), liderada pela Maria Fernanda Bacelar do Nascimento, que não era professora, mas sim investigadora, e que era a antítese da Mira Mateus — embora ambas viessem de algo acabado em UL (Universidade de Lisboa), eram grupos totalmente separados. Também já faleceu, mas era uma pessoa que tinha um trato muito bondoso e estava genuinamente interessada em partilhar os seus conhecimentos com engenheiros e informáticos, em vez de dividir para reinar. Infelizmente havia uma coisa em que nós divergimos completamente, que tinha a ver com a questão de tornar ou não os corpos públicos, o que levou a que a Linguatca acabasse por ser “competidora” do CLUL em vez de parceira. Algo de que tenho pena ainda hoje.

4. Algo chegou a surpreender você (de forma positiva ou negativa) no diálogo com os/as linguistas, ao longo dos anos de trabalho “conjunto”?

DS: Não sei por que falei apenas dos linguistas portugueses na pergunta anterior, mas devia talvez ter dito que eu, quando entrei para a IBM (em 1990, para o tal projeto de tradução automática), comecei logo a lidar com equipas internacionais que tinham informáticos e linguistas, e que aprendi tanto com uns como com outros. Além disso, havia pessoas que eram ambas as coisas, algo que desde que trabalhei em PLN foi coisa que quis ser, e que defendo que todos devemos ser. Uma das pessoas que mais me surpreendeu positivamente foi o meu colega finlandês Lauri Carlson (que mais tarde foi o meu orientador de doutoramento) e que era um especialista em linguística (e filosofia).

Ora numa discussão qualquer que estávamos a ter, ele retorquiu-me “mas isso é um simples *chart parser*” (uma tecnologia de análise de PLN), e daí descobri que ele tinha tirado o curso de informática por *hobby* na Universidade de Helsinque (e que, portanto, até sabia mais de tecnologias de análise sintática do que eu). E que me demonstrou que não interessa por onde se começa: uma formação em linguística seguida de uma formação em informática dá o “mesmo” resultado do que uma formação em informática seguida por uma formação em linguística. Quando eu digo “dá o mesmo resultado”, quero dizer para PLN, claro.

Outra coisa que me surpreende muitas vezes um pouco negativamente no diálogo com alguns linguistas é que estão à espera da minha ignorância — como é que uma pessoa que trabalha há vinte, vinte cinco anos em PLN continuaria ignorante sobre teorias e problemas linguísticos?

A outra coisa que me aconteceu muitas vezes e que me irrita sobremaneira (e muitas vezes eu até sei que é para isso), é colegas informáticos descreverem-me/apresentarem-me como linguista, e colegas linguistas apresentarem-me como “engenhaira”. Isso significa que me estão a reduzir ao outro lado, quando no PLN/LC não devia existir outro lado. Para mim, estamos todos do mesmo lado.

Finalmente, tenho de reconhecer com grande humildade e alegria que tive parcerias fantásticas na Linguateca com linguistas, que provam que não é nada do outro mundo.

5. Como chefe de laboratório/equipe, o que você diria para jovens linguistas que desejam se aventurar na área? Para você, qual seria o perfil ideal da pessoa de Letras que integra um grupo interdisciplinar de PLN?

DS: Eu diria: quanto mais cedo, melhor. Mas o mais importante é uma pessoa querer integrar uma equipe de PLN, estar interessada nos problemas que a equipe tenta resolver. E ter uma mente aberta. E confiar que aquilo que sabe pode ajudar a equipe a progredir. Ou que, pelo menos, se não sabe ainda precisamente esse ramo/essa teoria, recebeu na formação de Letras ferramentas para saber procurar/descobrir por si próprio ou própria, ou comunicar com a pessoa certa. Cada vez mais, na minha opinião, ser formado é saber aprender, saber procurar, saber comunicar, e saber ouvir.

Uma equipa é precisamente isso, pessoas diferentes que contribuem de maneiras diferentes com perspectivas diferentes e com histórias e personalidades diferentes. Se forem todos iguais, não é uma equipa, é um batalhão ;-)

6. Como você vê o PLN em português hoje, comparando com os anos iniciais? E o que você esperaria daqui para a frente?

DS: Eu não posso dizer que tenha uma visão muito conhecedora de como está o PLN do português hoje, nem sei se faz muito sentido comparar com os anos iniciais — ou melhor com os meus anos iniciais, porque não me arvorar em ser a primeira em Portugal de forma alguma. Isto porque a forma de fazer PLN (em qualquer língua) mudou bastante, e aliás há sempre a tendência pendular na ciência e na engenharia entre os métodos lógicos/declarativos e os empíricos/estatísticos, mas há duas coisas que, na minha opinião, e independentemente das modas, mudaram para pior:

(1) a tendência para as grandes companhias multinacionais como a Google, a Facebook, a Amazon liderarem (mesmo monopolizarem) a pesquisa, o desenvolvimento e até os locais (físicos, as nuvens, a computação de alto desempenho, HPC em inglês) onde se faz PLN. Isso é extremamente prejudicial

para uma pesquisa que tenha outros valores, culturais, nacionais, político-econômicos, como em princípio — ou pelo menos nos anos 1980, era o caso da pesquisa nas línguas nacionais.

(2) A segunda tendência é a da “internacionalização” da pesquisa, que significa que os governos nacionais em vez de desenvolverem textos e pesquisa nas suas línguas dão prioridade e incitam a que tudo seja feito em inglês: na Comunidade Europeia isso é terrível, a ponto de muita formação universitária dada em Portugal é agora dada em mau inglês. Não é a questão da soberania nacional que a mim me preocupa, é a ignorância total das consequências para um povo inteiro, que é privado de um direito que devia ser fundamental e que é o de pensar e raciocinar na sua própria língua, tornando-se um cidadão científico de terceira categoria ou estrangeirando-se necessariamente.

Eu sou uma prova provada de que é um perfeito disparate escrever em inglês para chegar a uma audiência “internacional”. Escrevi a minha tese de doutoramento em inglês sobre o contraste entre o inglês e o português, e estou convencida de que 99% dos seus (poucos) leitores tinham o português como língua materna e que, se eu a tivesse escrito em português, teria sido muito melhor escrita e teria sido lida por muito mais pessoas! Penso que no Brasil felizmente a situação não é tão má como em Portugal, que me desespera realmente. Mas considero que uma boa formação internacional começa por uma boa formação nacional em que cada povo desenvolve formas de comunicar, formas de pensar e formas de lidar com o ambiente (mesmo geológico e físico) que não são à partida globais. E que depois equipas internacionais com peritos de várias línguas/culturas/saber-fazer diferentes estão muito melhor equipados para lidar com desafios globais. Note-se, e insisto aqui que não me estou a referir exclusivamente nem especialmente às Letras — muito pelo contrário, a todos os domínios do saber.

O que é que eu espero daqui para a frente? Bem, vou antes dizer o que eu gostaria de que acontecesse daqui para a frente, em relação a cada um dos pontos: em primeiro lugar, que as pessoas tomassem consciência das consequências políticas de muitas opções e tecnologias que passam como “progresso”, e que parassem para pensar e para discutir. E que também apostassem no decrescimento tecnológico e em maneiras inteligentes de resolver problemas que não exijam supercomputadores.

Em segundo lugar, que houvesse uma maior consciência de que a diversidade é importante, que a diversidade linguística e cultural (assim como a diversidade biológica) é um bem, e que estudar e compreender a sua própria língua, assim como línguas de poucas pessoas, é tão importante e interessante como estudar os diferentes “íngleses”. E que, como falantes do português, deveríamos nos interessar por todas as variedades do português em vez de fragmentar em português brasileiro, português de Portugal etc. e talvez até apostar (também) num português internacional em vez de um *global English*.

7. Outras coisas que você queira dizer

DS: Acho que é muito interessante este tipo de entrevistas, porque todas as áreas, mesmo o PLN, têm já uma história (da ciência) que é importante conhecer para as pessoas que vêm trabalhar nela. Porque a ciência não é a-histórica, e que as teorias e contrateorias nascem num contexto intelectual e político. Um exemplo que demonstra isto muito bem é um dos meus livros preferidos, que também é sobre a história da ciência, neste caso da linguística e da filosofia, que é o de John M. Ellis, *Language, Thought and Logic*, em que ele faz uma síntese de três linguistas/filósofos: Wittgenstein, Pierce e Whorf. Mais do que contrastar e discutir as visões deles, ele explica porque é que cada um seguiu — em diálogo com outros — por caminhos diferentes.

Vi que, graças à pandemia, a ABRALIN fez entrevistas de vida muitos linguistas brasileiros, em particular à Charlotte Galves. Adorei a entrevista dela, a que assisti, e dou os meus parabéns à Cláudia por estas duas entrevistas (a propósito, estou ansiosa por ler a da Graça). Penso que seria muito interessante fazer para muito mais pessoas, tanto em Portugal como no Brasil, e que um livro de entrevistas poderia também suscitar mais interesse pela área. Até com redes de personagens, redes de contactos, redes de colaborações, de citações... Para uma história do PLN em português que não fosse só baseada no Google Scholar, no DBLP, e na ACL Anthology? Sem obviamente desmerecer destes projetos, seria interessante ter um conteúdo lusófono de quem conheceu quem, quem orientou quem, quem ouviu falar de quem... mais pessoal. Aqui fica uma ideia para um mais jovem, ou um grupo de mais jovens.

Finalmente, outra coisa que gostaria de dizer é que foi muito bom conhecer a Graça Nunes e o NILC (que, aliás, na altura era encabeçado também pela Lúcia Rino e pelo Bento Dias-da-Silva, este último um linguista) ao iniciar a Linguatca: foi uma parceria e um contacto luso-brasileiro muito positivo, que, mau grado todas as vicissitudes do meu percurso na Linguatca e fora dela, ainda agora dá os seus frutos, estando a organizar neste preciso momento um volume da *Language Resources and Evaluation* com o Thiago Pardo.

Se puder apelar a alguma coisa, é a de que continuemos a fazer PLN para o português, e não só para uma variedade do português; que mantenhamos um olhar internacional/lusófono sobre a nossa língua, e que com isso também ajudemos os nossos colegas em África, em Timor e em Macau a terem serviços e estudos sobre as suas realidades.

MINI GLOSSÁRIO

Abreu Faro

<https://tecnico.ulisboa.pt/pt/noticias/campus-e-comunidade/professor-manuel-abreu-faro-deixou-nos-ha-20-anos/>

“Recordado por um sem-número de alunos como um dos professores mais extraordinários do Técnico, capaz de lhes passar todo o entusiasmo que sentia pelo ensino das telecomunicações, o papel do Prof. Abreu Faro foi muito além das suas funções de docente, sendo indubitavelmente considerado como um dos motores de mudança do panorama científico nacional” (português).

Duas coisas que eu gostaria de realçar: a existência do complexo interdisciplinar do IST, lançada por ele, que foi criado no papel com 8000m², embora, no início, 7200m² fossem “futuras ampliações”, e que é um dos exemplos que para mim ilustram o provérbio “when there is a will there is a way”; e os variados livros de ensino escritos em português na área das telecomunicações.

ILTEC e Eurotra

Instituto de Linguística Teórica e Computacional, criado em 1988 para ser o parceiro português no projeto Eurotra, um projeto europeu muito ambicioso financiado pela Comissão Europeia de 1978 a 1992, cujo objetivo era a tradução automática entre sete ou nove línguas.

https://pt.wikipedia.org/wiki/Instituto_de_Lingu%C3%ADstica_Te%C3%B3rica_e_Computacional

<https://en.wikipedia.org/wiki/Eurotra>

INESC - Instituto de Engenharia de Sistemas e Computadores

É uma associação sem fins lucrativos criada em 1980 por alguns professores do Técnico (veja o verbete “Técnico”), nomeadamente José Tribolet e João Lourenço Fernandes, como um modelo de parceria indústria-universidade, inspirado no modelo americano, tendo como associados as principais universidades técnicas do país e as empresas de telecomunicações. O INESC congregou durante muitos anos grande parte da investigação nas áreas de engenharia eletrotécnica e informática do IST, além de ter tido forte intervenção no tecido empresarial português. Neste momento, existem 6 institutos que constituem o INESC.

<https://inesc.pt/pt/>

João Pavão Martins — agora professor catedrático de informática e fundador do curso de Engenharia informática no Técnico (veja o verbete “Técnico”), foi um dos primeiros doutorados em Inteligência artificial (IA) que voltou a Portugal, ao Técnico, sendo originalmente de engenharia mecânica. Além de académico, fundou uma empresa de IA com enorme sucesso, a SISCOG. Sempre muito preocupado com o ensino, além de criar uma terminologia portuguesa para a IA junto com os seus colegas de outras faculdades, escreveu vários livros de informática em português para ensino.

https://pt.wikipedia.org/wiki/Jo%C3%A3o_Pav%C3%A3o_Martins

Linguateca

É um projeto financiado pelo Ministério da Ciência e Tecnologia português de 1998 a 2011, e apoiado em termos informáticos pela FCCN até hoje. A sua missão é criar recursos para o processamento computacional da língua portuguesa e promover o paradigma da avaliação conjunta.

<http://www.linguateca.pt/>

Maria Fernanda Bacelar do Nascimento — seguiu a carreira de investigação (não de ensino) e pertenceu sempre ao Centro de Linguística da Universidade de Lisboa (CLUL), sendo ao tempo em que a conheci a líder do grupo dos corpos. Tinha trabalhado no Português Fundamental, um projeto feito à medida e tendo como exemplo “le Français Fondamental” e que, em Portugal, foi inicialmente liderado pelo professor Lindley Cintra. Na altura, tinha como principais colaboradoras/alunas a Luísa Alice Pereira e a Amália Mendes, e estava a desenvolver o Corpus de Referência do Português Contemporâneo (CRPC). A sua paixão eram os corpos orais. Ver para a sua biografia e bibliografia

https://pt.wikipedia.org/wiki/Maria_Bacelar_do_Nascimento

Mariano Gago

José Mariano Gago foi a pessoa que mais fez pela ciência em Portugal, e mais fez pela tecnologia linguística para o português, no mundo. Foi ele, como ministro da Ciência e da Tecnologia, que sugeriu que o processamento computacional da língua portuguesa fosse uma prioridade, discutida paralelamente à física, à matemática, à química, em 1999. E foi ele que permitiu e fomentou a existência da Linguateca.

Foi presidente da Associação de Estudantes do IST durante a ditadura e participou nas lutas estudantis contra o regime, tendo emigrado para a Suíça, onde participou ativamente na animação cultural para adultos na comunidade portuguesa. Doutorou-se em física de partículas e trabalhou no CERN até voltar a Portugal para ser professor do Técnico e depois se tornou o maior obreiro da política científica portuguesa durante décadas. Criou, entre muitas outras instituições, a Ciência Viva, que organiza uma conferência anual em sua honra.

https://pt.wikipedia.org/wiki/Jos%C3%A9_Mariano_Gago

Técnico - forma pela qual é/era conhecida a maior escola de engenharia portuguesa, de seu nome completo “Instituto Superior Técnico”, o “MIT português”, quando queremos apresentar no estrangeiro. Criado em 1910 para continuar escolas já existentes desde 1852, tinha inicialmente cinco ramos: Engenharia Civil, Engenharia Mecânica, Engenharia de Minas, Engenharia Electrotécnica e Engenharia Química (inicialmente Químico-Industrial), e esse perfil manteve-se até os anos 90. Fez parte da Universidade Técnica de Lisboa (UTL) de 1930 a 2013, altura em que a UTL e a Universidade (Clássica) de Lisboa se fundiram na Universidade de Lisboa. Desde a sua criação que fez parte da formação das elites portuguesas, como o demonstra a quantidade de ministros saídos das suas fileiras. Também a Associação de Estudantes do Instituto Superior Técnico teve um papel de relevo na contestação estudantil à ditadura do Estado Novo.

https://pt.wikipedia.org/wiki/Instituto_Superior_T%C3%A9cnico