

Számítógépes nyelvészet és mesterséges intelligencia –interjú Prószéky Gáborral

Prószéky Gábor 1981-ben az ELTE Természettudományi karán programtervező matematikusként, majd 1982-ben a Bölcsészkaron általános és alkalmazott nyelvészként végzett. 1994 óta a nyelvtudomány kandidátusa. A hetvenes évek végétől folytat szakirányú munkát, jelenleg az 1991-ben általa alapított, nemzetközileg is elismert MorphoLogic kutatócég ügyvezető igazgatója. A témakör első hazai szakkönyvének, a *Számítógépes nyelvészet (Természetes nyelvek használata a számítógépes rendszerekben)* című könyvnek a szerzője, a néhány éve megjelent *Számítógéppel emberi nyelven* címűnek társszerzője, valamint száznál több tudományos publikációnak az írója. Munkásságát a Neumann János Társaság 1995-ben Kalmár László-díjjal, 1997-ben a Worldcom kanadai kongresszusán pedig a World Young Business Achiever kitüntetéssel honorálták. 2000-ben a köztársasági elnöktől megkapta a Széchenyi-díjat, 2002-ben pedig az év informatikai menedzsere lett (Gyúrós Tibor-díj).

- Már a MorphoLogic előtt is sokirányú kutatásokat folytatott: bonyolult struktúrájú természetes nyelvek gépi reprezentációja, intelligens tudásbázis-lekérdezés, gépi fordítás...

- A MorphoLogicot megelőző időkben a saját kutatásaim inkább a hobbimnak számítottak. Amikor programtervező matematikusként már végeztem, de az általános nyelvészet szakon még nem, azt reméltem, hogy olyan állást találok, ami a

kettőt együtt engedi. A Nyelvtudományi Intézetben lett volna lehetőség, de jelezték, hogy még sincs. Ilyen típusú munkát máshol nem lehetett találni, így – mivel zenéltem közben – rögtön a Zenetudományi Intézetben kezdtem az életemet. A gépi nyelvészetet emellett csináltam magamban. Mindez öt évig tartott – elég sok anyagot gyűjtöttem, és mindenféle kisebb elméleti kutatást folytattam. Az akkori gépparkon nem lehetett komoly nyelvtechnológiai fejlesztéseket végezni. 1985-ben a Műszaki Kiadótól megbízást kaptam a *Számítógépes nyelvészet* című könyv megírására. De ez megint csak könyv volt, azaz nem szoftvert gyártottam. A kiadónak elfogyott a pénze, a kézirat nálam maradt, évekig házaltam vele, míg 1989-ben a Számalknál jelent meg. Megpróbáltam mindent leírni benne, lefedni a lényeges témákat, bemutatni a kapcsolódó irodalmat, és ehhez kapcsolatot teremteni mindenkivel a világban, aki számomra fontosnak tűnő kutatást folytatott.

A nyolcvanas évek végén az Országos Pedagógiai Könyvtár és Múzeumban dolgoztam, aminek volt egy nagyon jó kis számítástechnikai csoportja. Elég sok lehetőségem maradt a könyv befejezésére, illetve a máshol aligha megvalósítható kutatásokra. A könyv és a kutatások kapcsán kijutottam Hollandiába egy, az akkor még éppen csak induló európai közösségi projektek egyikén dolgozó kis céghez. Szerepe volt benne a nyelvtechnológiának. Konkrét rendszernek, intelligens tudásbázisnak a lekérdezése kapcsán érintette a természetesnyelv-feldolgozást. A 89 és 91 közötti projekt inkább azért volt érdekes, mert a későbbi webet, azaz az akkori lehetőségeket: a teletext, a videotex, a gopher és a többi hasonló technika egységes felületen történő kezelését szerette volna

megvalósítani. Ez volt az első eset, ahol a tudásomat nemcsak papírra kellett rakni, hanem valami készült belőle. Láttam egy európai projektet, másrészt az adminisztratív oldalát is, ami azért érdekes, mert hazajövelemtkor alapítottuk meg a MorphoLogicot.

- Hogyan indult a MorphoLogic?

- Nem akartunk feltétlenül céget alapítani, hanem az addigi saját tapasztalataimból összejött egy nyelvleíró rendszer váza, illetve nemcsak a váza, hanem a magyar nyelv szóalaktani leírása, és felmerült az igény, hogy mivel nem volt Magyarországon helyesírás-ellenőrző program, legyen ilyen. Ekkortájt váltott a Windows a 3.0-ból 3.1-re, és a Word szövegszerkesztő is megjelent Windows alatt. Ennek a 2.0 változatába került bele az első alkalmazásunk. Látszott, van igény ilyesmikre: a Word Perfect, vagy az XYwrite is magyarításra került. Kiderült, hogy a mi nyelvleíró formalizmusunk lényegesen hatékonyabb, mint amikkel mások kísérleteznek, és gyorsan sikerült magyar nyelven működő alkalmazásokat is létrehozni belőle.

Így indult a MorphoLogic. Az előzmények – a fióknak készített programok, a csak fióknak írt cikkek, a nemcsak fióknak írt cikkek, maga az elkészült könyvem és a kapcsolatrendszerem – mind elkezdtek élni. A világban akkor indultak a ma már nyelvtechnológiának nevezett alkalmazások. A nagy cégek ezekben járatos egyetemi, akadémiai emberektől kértek tanácsokat. Hihetetlen előnyt jelentett nekem, mert induláskor nem professzionális üzletemberekkel, hanem olyan gépi nyelvészekkel, azaz ugyanolyan „amatőr üzletemberekkel” álltam szemben, mint amilyen én magam is voltam, ez pedig

hihetetlen gyorsan működtette a lehetőségeket. Rájöttek, hogy ebben a régióban is kellene nyelvi megoldások – magyarra, lengyelre, csehre, románra – és, hogy van egy megbízható cég, amely ezt el tudja végezni. A Microsofttal például még gmkkorunkban kötöttünk szerződést. A világ is elég kezdetlegesen kezelte az ügyet. Ma már másként van.

Abban az időben csak az volt a kérdés, hogy meg tudjuk-e oldani a problémát, vagy nem. Korai indulásnál nagyon jó az ilyen.

Valószínűleg ez alapozta meg a cég működését. Illetve azok a fiókban lappangó, majd hirtelen életre kelő elvi dolgok. Ki voltak dolgozva, egyiket sem a hirtelen igény szülte. Ráadásul az induláskor – az embargó megszűntével – ugyanolyan gépparkkal rendelkezhattünk, mint a nyugati partnereink. A tapasztalat viszont sokkal inkább a bitek személyes ismeretén alapult, hisz mint tudjuk, nálunk nem voltak elkényeztetve a szoftveresek. Előnynek számított ugyanakkor az is, hogy a legjobbnak, legátgondoltabbnak tekinthető szoftvertechnológiát tudtuk indulásunktól alkalmazni. Ellentétben számos nyugati céggel, nem kötött minket semmilyen, tíz-tizenöt éve létező és folyamatosan javításra szoruló kód: úgy indultunk, ahogy akartunk. Tehát szoftvertechnológiailag, nyelvileg átgondolt, új elveken alapuló, a bitek miriádjai között magukat jól feltaláló profi programozókkal kezdtük.

A végére hagytam a legfontosabbat: a magyar nyelv problémáit kellett megoldanunk. Formai szempontból ugyanis az angol egyszerűbb a legtöbb világnyelvénél. Viszont, ha már megoldottuk a magyar nehéz problémáit, akkor olyan ötletek és megoldások vannak a kezünkben, amik nem jutnának egy angolnak, vagy amerikaiak az eszébe. Nem azért van ez, mert szeretjük azt

hangoztatni, hogy „mi magyarnak születtünk, és más van a fejünkben”, hanem mert itt tényleg a nyelvről volt szó. Bonyolult a nyelvünk, a problémáit viszont ugyanakkora helyen, ugyanannyi időben kellett megoldanunk. Amikor megvolt a szóalaktani leírás, az alkalmazásokat már könnyebben le lehetett belőle vezetni, sőt, a tapasztalataink segítettek más nyelvekre is hamar megoldani a hasonló problémákat. Egy példa: a lengyel és cseh nyugati szemmel nézve bonyolult. A mi oldalunkról nézve, a magyar után a leírást csak egyszerűsíteni kellett: így vált előnnyé a magyar nyelv. A mai napig előszeretettel mondjuk, hogy ha valami jól működik a magyarra, akkor a többire se okozhat nagy gondot. Nem állítom, hogy a világ összes nyelvét ebbe a kategóriába teszem, de maradjunk az Unióban... Ez azért lényeges, mert mi valóban egyazon formalizmussal estünk neki a különböző nyelveknek. Amikor létrejött a szóalaktani, később mondattani leírás, az alkalmazások teljesen függetlenek voltak a nyelvtől. Különböző alkalmazások használják a nyelvi leírásokat. Különböző platformokon, különböző gépeken és operációs rendszereken működnek. A világban sokszor nem létezik egységes platform, csak egységesnek tűnő marketing: egy-egy nagy cég A-tól ezt, B-től azt, C-től pedig amazt vette meg, majd mindegyikre ráragasztotta, hogy az övé. De ezek az eljárások belül teljesen inkompatibilisek. Viszont amikor nekünk hirtelen és ügyesen kell váltani, nagy előny, ha egy ilyen közös platform létezik, és tényleg ahhoz tartoznak a nyelvek és az alkalmazások.

- Hogyan ítéli meg az MI és a gépi nyelvészet viszonyát?

- Az MI minden intelligensnek gondolt tevékenységet lefed, így a nyelvfeldolgozásét is. Viszont a számítógépes nyelvészet, a

nyelvfeldolgozás elég hamar „önállósodott”. Chomsky az 50-es években zseniálisan kitalált valamit, ami a formális nyelvek elméletében óriási jelentőségű, aminek gyakorlatilag az összes programozási nyelv és elemezhetőségük a létét köszönheti. Ám az emberi nyelvek esetében gyorsan rájöttek a számítógépes kutatók, hogy a Chomsky-modellek saját maguk demonstrálására ugyan megfelelőek, de a nyelv jó működtetéséhez rengeteg további dologra van szükség. Hamar elvált a számítógépes nyelvészet paradigmája Chomskytól. Elméleti nyelvészeti oldalról viszont nem létezett alternatíva: mindegyikbe „belelógott” a pszichológia, az MI, stb. A hatvanas és a hetvenes években ezért haladt együtt az MI és a nyelvészet. A hetvenes évek végén azonban megjelentek az alternatív generatív – számítógépen is jól kezelhető – nyelvelméletek, és belőlük lett modern generatív számítógépes nyelvészet, míg korábban a számítógépeseknek, illetve az MI-kutatóknak az egyedi, részben ad hoc, részben más, az ember egyéb tevékenységeit szintén lefedő kitalációja működött. Gyakorlatilag ekkor mozdult el a gépi nyelvészet a kimondottan a nyelv leírásával foglalkozó paradigmáktól. Ugyanakkor szintén a nyolcvanas években jelent meg az első Macintosh, majd a PC-k, aztán a rajtuk futtatható szövegszerkesztők, melyeknek a fejlesztésébe a kilencvenes évek elején már mi is be tudtunk kapcsolódni. Ezeken az alkalmazási területeken gyors, hatékony eredményt kellett elérni. Az ilyen jellegű feladatokra kezdtek koncentrálni a gép nyelvészek, és a nyelv emberi kommunikációban, pszichében betöltött általános szerepe egyre kevésbé tűnt lényegesnek az alkalmazások világához képest. Majd egyszer. Ha olyan okosak lesznek a rendszereink, hogy túljutottak a szavakon, mondatokon,

szövegegységeken, akkor megint előjön az MI, és ennek bizonyos jeleit körülbelül mostanában láthatjuk. Vannak külön-külön jól működő nyelvtechnológiai modulok, de az egészet az ember mozgatja. Chomsky gondolatára visszatérve: még soha nem láttunk olyat korábban, hogy a nyelv ki lenne szakítva az emberből. Ő pedig már az ötvenes években azt ígérte, hogy a nyelv önmagában, minden jelentése és a kommunikációs helyzet említése nélkül leírható. Az a szint a számítógépesek szerint, ameddig leírható, szinte mindig találkozik a másikkal, a világgal. A gépi rendszerek – mivel alkalmazások, és az alkalmazásoknak céljuk van – elég jól, hatékonyan kezelhetők, hiszen gépeink kapacitása megnőtt. Tehát a szövegek száma folyamatosan nő, és a gépeknek nem csak a helyesírást kell ellenőrizni, hanem intelligens lekérdezéssel, vagy az ismét előkerült gépi fordítással kell foglalkozniuk. Ezek olyan magas szintű emberi produkciók, melyekben ugyan benne van a nyelv, de nem mondhatjuk azt, hogy egymagában van benne, és a humán információkezelés többi intelligens részrendszere nem számít. Azt szoktam mondani, hogy manapság itt van az MI és a számítógépes nyelvfeldolgozás újbóli összeházasodásának az ideje. Ha nagyon általános dolgokat – számítógépes látás, beszédfelismerés, karakterfelismerés – nézünk, akkor viszont mindig található kapcsolódási pont. Ráadásul a beszédfeldolgozást idesorolják, ugyanis a nemzetközi terminológia *language and speech*-ről beszél, ami nem áll, csak esetleg *text and speech*. A gépi világban a *language* igazán a leírt és nem az ember számára elsődleges írott nyelv. Azt szoktam mondani, hogy az írott nyelv a számítógép anyanyelve. Ha beszélt szöveget adok a gépnek, azt is előbb-utóbb digitális szekvenciává konvertálja át, ami pedig

már egy karaktersorozat, azaz olyan, mint az írott nyelv. Ezzel szemben az igazi beszélt nyelv az embernek elsődleges, míg a gépnek másodlagos, és fordítva. Mindezt azért mondom, mert – nem mintha ki szeretném zárni a beszédet – most csak a mi kutatásaink kapcsán elsősorban az írottról beszélek.

- Miként látja a MorphoLogic szerepét a hazai gépi nyelvészetben?

- Tizenhárom éve spontán kaptunk egy-két olyan szerepet, melyekre nem voltunk predesztinálva, de így alakult az élet. Elmondható a helyesírás-ellenőrzőről is, hogy lehet jobbat csinálni, de ez úgy jött létre, hogy a megbízók elégedettek voltak vele. A felhasználók – sokszor teljes joggal – nem annyira elégedettek. Kérdés, kapunk-e arra egy nemzetközi magáncégtől megbízást, hogy még jobb legyen a magyar helyesírási program. Általában nem, mert a piacot nézik, és ha nincs több eladás, megállapítják: nem kell semmi, még akkor se, ha jobb.

A magyar nyelv tökéletes számítógépes leírása tudományos akadémiai típusú feladat, amire állandóan megvan az igény bennünk. Viszont már létező, magánkézből lévő termékrendszer javíthatására nehéz támogatást kapni, még ha az a termékrendszer kiemelt feladatot lát is el, a mi esetünkben „irányítja” a – naponta többszáz ezer – gépelő helyesírását. Más tehát egy prototípussal végződő kutatás-fejlesztési feladat, és más egy piacon lévő termék javítása. Ezért a MorphoLogic egész életében létezett ez a kettősség: kutatóhelyről jöttünk, akadémiai típusú emberek vagyunk, és akkor is végeznénk a dolgunkat, ha elfogyna a célfeladatra adott pénz, viszont egy magáncégben ez mégsem mehet, hisz akkor tönkremennénk. A MorphoLogic a hazai

piacon szokatlan módon tölti be ezt a kettős, akadémiai/piaci szerepet: évekig csak magunk finanszíroztuk a saját alap kutatásainkat is. Aztán két-három éve elindultak a nyelvtechnológiát fontosnak tartó kutatásfejlesztési projektek, és állami támogatás is van rájuk. Nagyon jó, hogy ezekben vannak partnereink. Most már más intézményekben is létrejöttek, vagy kezdenek létrejönni a téma iránt érdeklődő csoportosulások.

Tavaly decemberben volt Szegeden az első magyar számítógépes nyelvészeti konferencia száz résztvevővel, ami azt jelzi, hogy van érdeklődés. Remélem, a csoportokkal és konzorciumokkal való együttműködéssel megvalósulhatnak olyan dolgok, melyeket a csak MorphoLogicból álló magyarországi számítógépes nyelvészet korábban formai nehézségek miatt nem tudott megoldani.

Az egyik most befejezett projektünk eredménye, hogy üzleti szövegek lekérdezhetőségét tettük lehetővé. A szövegből kivonatoltuk a tulajdonképpeni üzleti eseményt: mi történt, eladták, megvették, tönkrement, ki, mennyiért, stb. Rengeteg esemény van kódolva egy szövegben, és ha rákeresünk, azért nehéz az effajta tartalmat megtalálni, mert mindig máshogy fogalmazzuk meg. Még nem létezett ilyen szövegkivonatoló rendszer, és ezt nem lehet megcsinálni termékként. Egyelőre csak prototípus. Ugyanakkor elmondható, hogy korlátlan a MorphoLogic, vagy a nyelvtechnológia jövője. Nő, egyre nagyobb igényeket támaszt a terep. A szövegek nem kezelhetők hagyományos módon. Nagyon igyekszünk, de mit tud például egy kereső manapság? Se fogalomra, se tartalomra nem képes keresni, hanem csak betűsorra. Az emberi keresés sokkal komplexebb, és ha valamit átveszünk belőle, talán javíthatunk a

találathalmaz pontosságán. Más a pontosság, és más a lefedettség. Egyre jobban elérhetőek a szövegek, azaz egyre jobb keresők, fordítók, létrehozás-támogató eszközök kellenek hozzá.

- Milyen projekteken dolgoznak jelenleg?

- Nagyon sokon... Azzal a technológiával, amivel a magyarnak „nekiestünk”, majd más nyelvekre alkalmaztunk, kis uráli rokon nyelveket írnak le a magyarországi uralisztikai tanszékek.

Az üzleti szövegkivonatoló-rendszerhez lekérdezőt építünk, és a hatékonyabb használaton dolgozunk.

A már lefordított szövegek és az eredetijük összevetésével terminológiákat gyűjtő kutatás is folyik. Sokszor fordult elő, hogy ugyan lefordítottak szövegeket, viszont nem kristályosodott ki a terminológia. Ez nagyon lényeges; a jogharmonizációs szövegek fordításakor is láttuk. A gép maga ajánlja majd fel az eredetiből és a fordításából kigyűjtött terminológia-párokat.

Az összes számítógéppel és nyelvvel foglalkozó kutatónak van egy vesszőparipája, ami általában sokszor úgy jelenik meg, hogy a fordítás megoldhatatlan probléma. Szerintem is az, ugyanakkor mindenkinek egy régi álma ez. Tulajdonképpen a számítógép kialakulásában is jelentős része volt annak, hogy gyorsan kellene fordítani. Tehát, nagyon régi, állandóan előjövő témáról van szó. Ugyanakkor szögezzük le: az ember tudását ez ügyben sem tudja olyan mértékben átadni a gépnek, mint ahogy szeretné. Mert akármilyen hatékony a gép – ez az egyetlen, amiben többet tud nálunk –, képtelen elsajátítani azt a háttérrel, hogy „kétéves vagyok csak, de ember vagyok”. Azaz: még alig tudok valamit az anyanyelvemen, viszont korlátlan számú megnyilvánulást hallok, amiket mind építek be, és egy, két, három nap alatt sokkal többet

tudok tenni velük, mint bármilyen gép. Egy gépbe mindenfélét beletehetek két év alatt, csakhogy az továbbra sem emberből van. Az igény viszont óriási, mert például az internetező magyar embernek – az Unióba belépve – angolul kell meglátogatnia egy rakás portált. Egyet-kettőt lefordítottak ugyan, de az összes többin is ott a támogatásokra vonatkozó információ. Ezeket már nem fordítják le, mert nincs rá se mód, se idő. Tehát a másik nyelvű információ olyan fajta megértését lehetővé tevő eszköz kell, hogy rájöhessenek, ez nekem kell-e, vagy sem, és tényleg le kell-e fordítanom. Bár lehetetlen, de szükséges az ilyen eszköz. A kísérletek két szélsőséges kategóriába esnek: a nyelvnek mindenfajta részletével foglalkozó, a mennyiséggel nem törődő, nyelvi problémamegoldás iránt érdeklődő, azt tíz mintamondattal megoldó akadémiai jellegűek, illetve a „fogd a pénzt, és fuss”-típusúak. A kettő közötti optimumot elég nehéz megtalálni. Melyik az az eszköz, ami már elég sokat tud, és mégis eladható, mert igény van rá? Ugyanakkor ez messze nem olyan, amilyen kellene, hogy legyen. De hol van az az optimum, ahol még az eladhatóság és már a tudomány így ér össze? Ez a szép ebben a feladatban. Megoldani nem lehet, de lehet egyfajta optimumra törekedni. Azon dolgozunk, hogy a közeljövőben, még 2004-ben megjelenjen egy olyan eszköz, ami tesz valamit ebbe az irányba. Ha odamegyek egy angol szöveghez, villantson fel valamit magyarul abból a tartalomból, ami ott van, és amiről azt gondolom, hogy többet tudok!

Az utóbbi évek egyik legsikeresebb általunk készített programja a MoBiMouse volt. „Egyérintős” módszer: otthagynom az egeret, ő meg azonosítja a képernyőn a szöveget, és valamilyen szótári információt rendel a szóhoz, vagy – kontextus alapján – a

kifejezéshez. Most ezt a módszert általánosítjuk, és az új változat a mondatról mond valamit. Azaz: a mondat le lesz valahogy fordítva. Nyilván messze lesz a tökéletességtől, de azt már el lehet várni tőle, hogy orientáljon, mi is van itt a képernyőn. Nem az ember helyettesítése a cél, mert a fordító azért profi, mert úgy fordítja a szöveget, ahogy más nem, és a számítógép még annyira sem. Az új program nem ülteti át tökéletesen a szövegeket, viszont azonnali segítségként jelentkezik. A MoBiMouse logikája után ez a MoBiCAT, tehát a MorphoLogic Bilingual CAT (*Computer-Aided Translation, számítógéppel segített fordítás*). A MoBiDic után a bálnát „levittük egérbe”, így lett MoBiMouse, most meg „felment macskába”...

A MoBiCAT rendszer az általunk évek óta készített MetaMorpho fordítórendszernek az első konkrét megnyilvánulása. Tíz-tizenegy évig építettük a különböző modulokat, majd összeszedtük valamennyit, és három éve kiindulási alapként elindítottuk ezt a fordítóprojektet. Most érte el azt az állapotot, hogy nagyon hatékonyan javítható – és ez igen fontos. Megvan az alaptudása, és ezért hamar megjelenhet az első rá épülő eszköz. Nyilvánvalóan segítő lehet az internetes böngészésben azoknak, akiknek gondjuk akad az angollal. Állandóan javul a minősége, és remélni lehet, hogy másfajta alkalmazásokban is meg fog jelenni. És nemcsak angol-magyar irányban, de először csak abban. A magyar-angolra már előbb említett akadémiai/egyetemi társainkkal, az MTA Nyelvtudományi Intézetével és a Szegedi Egyetemmel konzorciumként beadtunk egy pályázatot. Az angol szövegeket ugyanis a magyar beszélő

akarja megérteni, de amikor magyar a szöveg, ahhoz nem megértésre, hanem teljes szövegfordításra van igazi igény.

A MoBiCAT-tel egyidőben wapos formában is meg fog jelenni, hiszen fordítónk így is működik, és az ilyen jellegű MoBiWAP-szótárszolgáltatásunk eddig is elég nagy érdeklődésre tarthatott számot.

Azt a munkát kell a géppel hatékonyabbá tenni, amit a gép tud jól csinálni, és meghagyni az embernek, amit az ember tudhat csak. Egy másik cél pedig továbbra is az, hogy kíváncsiskodunk, hogyan működik az ember, tehát nem adjuk fel az MI igazi, eredeti célját, hogy rájövünk, miként működteti az ember a nyelvet, és milyen fajta interakciók alakítják ki azt, amit nyelvprodukciónak, vagy nyelvmegértésnek tekintünk. Ezeknek az egyszerűbb vetületei a való életben felhasználhatók. Attól, hogy valami tudományos, nem biztos, hogy nem használható hatékonyan, és ha valami jól működik, annak bizony sokszor nem a hackeléssel történt megoldás az oka, hanem lehet tudományos alapja is. Ezt a kettőt próbáljuk összehozni, ez az ars poeticánk.

- Miben látja a kutatásfejlesztési projektek sikerének, eredményességének titkát?

- Időben kell észrevenni és megfogalmazni a problémát, aztán gyorsan elindulni a megoldás irányába. Magyarul: akkor van nagy esélye egy kutatás-fejlesztési projekt sikerességének,

(1) amikor már van olyan alapgondolatunk, aminek a mentén el tudunk indulni, és

(2) amikor ezzel egy időben a lehetséges konkurensok még csak azt mérlegelik, hogy érdemes-e egyáltalán a témával foglalkozni.

- Ha visszamehetnénk az időben, és most lenne egyetemista, mivel foglalkozna legszívesebben? Milyen témakörben, kutatási területben látna komoly perspektívát?

- Biztosan állítható, hogy ma fontosabb területnek számít a számítógépes nyelvfeldolgozás, mint amikor én kezdtem a kutatásokat a hetvenes évek végén. A Pázmány Péter Katolikus Egyetem Információs Technológiai Karán a nyelvtechnológiát főként oktatva nyugodtan kijelenthetem, hogy a téma iránt érdeklődő egyetemi és doktorandusz hallgatóim száma és egyre többük elkötelezett érdeklődése egyértelműen ezt bizonyítja. A hatékonyan kezelhető szövegek méretének hihetetlen növekedése és az internet megjelenése a gépi nyelvészet számára óriási lehetőségeket, hatékonyan művelhető alap kutatási területet és rengeteg olyan fontos alkalmazást biztosít, aminek a megvalósítása tíz-húsz éve még képtelenségnek tűnt. Így tehát biztosan állíthatom, hogy ha ma lennék egyetemista, ugyanezt a kutatási területet választanám. A problémák egy része ma sem megoldott, így például a gépi fordítás ugyanazokat a nehézségeket rejti, mint húsz vagy harminc éve, csak akinek ma jó megérzései vannak, az gyorsabban tudja ezeknek az intuícióknak a helyességét vagy helytelenségét igazolni, így a húsz-huszonöt évvel ezelőttinél lényegesen hatékonyabban lehet ma végezni a kutatásokat.

- Milyen alapelveket érdemes képviselni ahhoz, hogy a csúcstechnológiai kutatásokban komoly eredményeket érjünk el?

- Ahhoz, hogy eredményesek legyünk, tudni kell, hogy mi számít eredménynek a világban, azaz ismerni kell a területen folyó

kutatás-fejlesztéseket, olvasni kell a szakirodalmat. Ugyanakkor nem szabad azzal túl sokat sem foglalkozni, hogy mások mit csinálnak, nehogy a saját intuíciónk lássa ennek kárát. Egy más területről vett példával azt szoktam mondani, hogy amelyik zenész úgy akar megtanulni improvizálni, hogy állandóan Bach improvizációként született műveit tanulmányozza tudományos alapossággal, az sokszor már nem is mer rögtönözni, mert fél, hogy Bach ezt is, meg azt is jobban csinálta... Ebben kétségtelenül van némi ellentmondás: egyrészt figyelni kell a többiekre, másrészt nem szabad a többiekre figyelni. Aki itt megtalálja az egyensúlyt, annak van esélye, hogy a csúcstechnológiai területeken is nagyot alkothasson.

<http://www.morphologic.hu> (MorphoLogic),

http://www.morphologic.hu/h_pg.htm (Prószéky Gábor)