

ZÁRT KORPUSZOK ÉS PÓTKOMPETENCIA

FORGÁCS TAMÁS

Külföldi szakirodalmi munkákban egyre többször lehet az utóbbi időben találkozni olyan próbálkozásokkal, amelyek a modern nyelvészeti irányzatok vizsgálati módszereit próbálják felhasználni a diakrón kutatásokban. Ebben szerepe lehet annak az útkeresésnek is, amely a modern elméletek – főként a generatív grammatika és annak továbbfejlesztései – korlátainak felismerését és a beljük vetett hit megrendülését követte (vö. Behrens 1989: 1, Szemerényi 1992: 11 kk.). Így próbálkozások történtek a generatív modell diakrón alkalmazására is (pl. King 1971 és Van de Velde 1971), de még inkább egyes strukturalista irányzatoknak, főként a valenciaelméletnek a történeti szövegek kutatásában való felhasználására (pl. Korhonen 1978, Greule 1982, Greule 1982a, Ágel 1988).

A történeti szövegek szinkrón írásbeli nyilatkozatok (utterance) véges számú halmazának, azaz korpusznak tekinthetők. A korpusz-fogalom bevezetésében és elméleti megalapozásában az amerikai strukturalizmus bloomfieldi irányzatának, ezen belül is főként Harrisnek vannak múlhatatlan érdemei. Noam Chomsky és követői ugyan elvetették a korpuszkutatást, mondván hogy a korpuszok szelektívek és akcideniálisak, ennél fogva alkalmatlanok az anyanyelvi beszélő-hallgató kompetenciájának mérésére és demonstrálására, mégis az utóbbi időben az figyelhető meg, hogy a korpuszok problematikája a nyelvészeti módszertani viták homlokterébe kerül. Elsősorban a szociolingvisztikai és a pragmatikai kutatások körében tapasztalható erős ellenállás azzal az irányzattal szemben, amely az intuíciót teszi meg a nyelvészeti kutatások egyedüli bázisának, azonban nyilvánvaló, hogy a nyelvtörténeti kutatások is a nyelvleírásnak azokhoz a területeihez tartoznak, ahol a kutató korpusz nélkül egész egyszerűen nem tud létezni. Ugyanakkor azonban – éppen a régi korok szövegeiben található nyilatkozatok véges száma miatt – elkerülhetetlen, hogy megvizsgáljuk ezeknek az elméleti státuszát a korpusz-elméletben és ezáltal tisztázzuk, mennyire alkalmasak régi korok nyelvállapotainak, azaz tkp. „holt nyelveknek” a leírására.

Ha a nyelvészeti elemzés egy korpuszon alapszik, akkor ez azt jelenti, hogy a kutató rendelkezik egy véges hosszúságú szöveggel, amelynek elsősorban két követelménynek kell megfelelnie: álljon lehetőleg nagyszámú nyilatkozatból, amelyek a vizsgált nyelvben valóban elhangzottak. (A textuális összefüggés nem feltétlen követelmény.) Coseriu (1975: 39) szerint feltételezhető, „hogy egy viszonylag nagy korpuszban már majdnem minden vagy minden megtalálható, ami a nyelvre jellemző [daß in einem ziemlich großen Corpus fast alles oder alles, was für die Sprache maßgebend ist, zu finden ist]”. A generativisták szemszögéből különös hangsúlyt kap azonban a korpuszok véges volta. Így Sitta 1971: 91 megjegyzi, hogy hosszabb korpuszok esetében is számolni kell azzal, hogy „nem minden adatolható, ami lehetséges [nicht alles, was

möglich ist, auch belegt ist]”. Ennek olyan veszélyei lehetnek, hogy a rendszer egyes fontos területei nem is kerülnek be a vizsgálatba, illetve az adatolt, ennél fogva vizsgált területeken belül az adatok nem a tényleges viszonyokat tükrözik. Éppen ezért elkerülhetetlen, hogy ilyen vizsgálatok esetében átlépjük a korpuszok határát. Ez élő nyelvekből származó korpuszok esetében informátorok, azaz a vizsgált nyelvet kompetensen beszélő személyek bevonásával, kikérdezésével történhet.

A vizsgálatot végző kutató, az informátorok és a korpusz viszonyában Greule (1982: 71) három esetet különböztet meg:

1/ A kutató egyáltalán nem ismeri a vizsgált korpusz nyelvét. Ebben a – nyilvánvalóan ritka – esetben kezdettől fogva informátorokra van utalva, egyébként a vizsgálat a rejtvényfejtés jellegét ölti (gondoljunk csak régi feliratok, pl. egyiptomi hieroglifák megfejtésére).

2/ A kutató ismeri a vizsgált nyelvet, de maga nem kompetens beszélője (kompetenter Sprecher) annak. Ebben az esetben önállóan is képes a vizsgálatot végezni, ha azonban valamiben bizonytalan, ki tudja magát segíteni informátorok megkérdezése révén.

3/ A kutató maga is rendelkezik az adott nyelvi kompetenciával. Ez az eset a leggyakoribb a mai szinkron leírások készítése során. Ebben az esetben a vizsgálatot végző személy tkp. önmaga informátora, bár korrektebb és objektívebb a vizsgálat, ha csupán első, de nem az egyetlen informátora.

A fenti három eset azonban csak élő nyelvek vizsgálata során lehetséges. Élő nyelvek korpuszait ezért Greule (1982: 72) „nyitott” korpuszoknak (offene Corpora) nevezi, ezekhez ugyanis elérhetőek informátorok. „Holt” nyelvek korpuszainak, azaz a régibb korok nyelvi állapotát dokumentáló szövegeknek ezzel szemben a „zárt” korpusz (geschlossenes Corpus) elnevezést adja (uo.). Ezek vizsgálata során sem informátorok nem állnak rendelkezésünkre, sem az elemzést végző kutató nem rendelkezik a korpusz nyelvi állapotának megfelelő kompetenciával. Ha azonban a mai magyar nyelvi kompetenciával rendelkező személy vizsgálja pl. a Müncheni Kódex szövegét, akkor feltételezhetjük, hogy legalábbis a fentebb 2/ pont alatt említett eset fennáll, azaz az elemzést végző kutató a középmagyar kor nyelvének és a mai magyar nyelvnek közvetlen (genetikus) rokonsága alapján a korpusz nyelvét legalábbis „ismeri”, anélkül, hogy azt kompetensen beszélné. Ez az ismeret ugyan nem elég a korpuszbeli nyelv teljes grammatikai leírásához, ugyanakkor viszont kellő alapot szolgáltat a vizsgálat megkezdéséhez.

Ahhoz azonban, hogy zárt korpuszok nyelvi leírása – ha nem is teljesen, de legalábbis megközelítően – olyan megbízható legyen, mint a nyitott korpuszoké, a vizsgálatot végző személynek törekednie kell arra, hogy minél közelebb kerüljön a korábban 3/ pont alatt jelzett állapothoz, amelyben nincs szükség informátorok bevo-

nására, mert ő maga is – csaknem – rendelkezik a korpuszbeli nyelvet illető kompetenciával. A nyelvész tehát rákényszerül arra, hogy kiépítse a saját „pótkompetenciáját” (Ersatzkompetenz). Ez ugyan nem éri el egészen a szöveg szerzőjének (aki „native speaker”) a kompetenciáját, ám a hosszan – több éven át – tartó, beható foglalkozás a korpusz nyelvvel lehetővé teszi egyfajta „secondary native intuition” elsajátítását (Closs 1969: 369, idézi Ágel 1988: 18).

Hasonlóképpen fogalmaz Roger G. van de Velde is, aki behatóbban foglalkozik azzal a kérdéssel, miként lehet régibb korok nyelvi állapotait generatív-transzformációs alapon leírni. Szerinte az ilyen vizsgálatokban a szöveg szerzőinek ill. leíróinak ideális nyelvi kompetenciája (idealer Sprachkompetenz der Textschreiber) helyébe – kényszerűségből – a vizsgálatot végző nyelvész kompetenciáját kell léptetnünk. (Ez azonban nem lehet annak mai nyelvi kompetenciája, nagyon kell vigyáznunk arra, hogy ne vetítsük vissza mai nyelvi kompetenciánkat régibb korok nyelvi állapotára.)

Van de Velde 1971: 40 k. szerint a kutató (der Deskribent) kompetenciáján belül négy kompetencia-típust kell elkülöníteni:

1/ *megértési kompetencia* (Verstehenskompetenz), azaz a vizsgált nyelvi anyag(ok) szemantikai-funkcionális megértése;

2/ *szövegkompetencia* (Textkompetenz), azaz a külső és belső szövegösszefüggés figyelembevétele;

3/ *filológiai kompetencia* (philologische Kompetenz), azaz a megértéshez szükséges paleográfiai, kultúr-, vallás- stb. történeti ismeretek, végül

4/ *pótkompetencia* (Ersatzkompetenz), azaz a nyelvi anyaggal való beható foglalkozás révén kialakított, másodlagos kompetencia.

Ezek közül az első kettőt tkp. összevonhatjuk, hiszen a megértést a szövegösszefüggés figyelembevételétől elég nehéz lenne elválasztani. Az így egyesített megértési kompetencia pedig a filológiai kompetenciával együtt inkább csak a pótkompetencia előfeltételének, korábbi fázisának, mintsem azzal egyenrangú típusnak tekinthető (vö. Greule 1982: 74. k.).

A régi szövegekkel való legbehatóbb foglalkozás, a legelmélyültebb filológiai felkészültség mellett is marad bizonyos hátránya a nyelvésznek a szöveg szerzőjével szemben: az ő (pót)kompetenciája nem kreatív, azaz nem képes végtelen számú új mondatot előállítani és megérteni. Igaz ugyan, hogy kísérletképpen ő is alkothat ilyeneket, de grammatikalitásukat nem tudja igazán megítélni, legfeljebb a korpuszokban előforduló példák alapján valószínűsítheti, hogy azok jó vagy rossz mondatok-e. Ezért a generatív elmélet alapján álló leírások számára tkp. még a pótkompetencia sem képes igazán adekvát bázist teremteni: „[...] der Beschreibende (hat) damit zu rechnen, daß es ihm nie möglich sein wird, die ihm nicht als Muttersprache vertraute Faktizität älterer Sprachstufen in Grammatikalitätsgraden oder in Akzeptabilitätsstufen so zu

charakterisieren, wie es ihm nach dem Ideal der transformationell-generativen Grammatik auf Grund seines Sprachvermögens für seine Muttersprache gelingen würde. [...] Auch im Bereich der Sprach- und Geistesbildung, der Welterschließung durch die (Mutter-)Sprache, der sprachlichen regelgebundenen Kreativität ist es nicht möglich, eine ältere Sprache in der *gleichen Weise* zu erfassen, wie sie von den ursprünglichen „native speakers“ beherrscht und verwendet wurde“ [a leírást végző személynek számolnia kell azzal, hogy sohasem fogja tudni régibb korok nyelvi tényeit grammatikalitási és akceptabilitási fokok szerint úgy jellemezni, mint ez számára a transzformációs-generatív nyelvtan ideálja szerint nyelvi tudása alapján anyanyelvét illetően sikerülne. [...] A nyelv- és gondolatalkítás, a világnak az (anya-)nyelven keresztül történő szemlélete és a nyelvi szabályszerűségekhez kötött kreativitás tekintetében sem lehetséges egy régibb nyelvet ugyanazon a módon feltárni, mint ahogyan azt az igazi „native speaker“-nek tudták és alkalmazták] (Van de Velde 1971: 49).

Egy strukturalista indíttatású vizsgálat azonban nem foglalkozik mondatok generálásával, csak a meglévő mondatok leírásával. Ebben az esetben – kiindulva a szöveg szerzőjének ideális kompetenciájából – minden előforduló szerkezetet és mondatot szükségszerűen grammatikálisnak fogadunk el (vö. Van de Velde 1971: 49, Greule 1982: 74).¹ A pótkompetenciának tehát már nem kell foglalkoznia a grammatikalitás kérdésével, feladata a megértésre és az ezzel összefüggő metanyelvi funkciókra (nyelvi struktúrák felismerése, helyes szegmentálása, nyelvi elemek helyes klasszifikálása stb.) korlátozódik. Ezeknek a kívánalmaknak azonban a kellően kidolgozott pótkompetencia túlnyomórészt eleget is tud tenni.

A pótkompetenciát azonban – mint említettük – először ki kell alakítania a kutatónak. Ez nem jelent mást, minthogy a lehető legjobban le kell dolgoznia azt a nyelvi és kulturális „deficitet“, ami a vizsgált anyagtól elválasztja. A pótkompetencia kialakítása hasonlít némiképp az idegennyelv-tanulásra (Zweitspracherwerb), de természetesen megkönnyíti a folyamatot, ha a „holt“ nyelv – mint a fent említett esetben is – a leírást végző személy anyanyelvének egy korábbi állapotát rögzíti, annak mintegy „történelmi variánsa“, mondhatnánk „historiolektusa“ (vö. Greule 1982: 75). (Ugyanakkor azonban ez egyben az az eset is, amikor a lehető legnagyobb a veszélye annak, hogy a deskribens a saját anyanyelvi kompetenciája alapján nyert intuícióját rávetíti az adott diakrón nyelvállapotra, s ez esetleg hibás következtetésekhez vezet. Ennek elkerülése, s a kellő objektivitás megtartása szinte egyfajta „parancsolat“ kell legyen a vizsgálatot végző személy számára.)

¹ Kivéve a nyilvánvaló tévedések, másolási hibák eseteit. Pl. MünchK. 49 va, Márk XIII, 1: *Mefte2 tē kenLēd mēl' kouēc v' mel' raka* [oc. Hogy itt nem nominális állítmánnyal, hanem a *tekint* 'néz' ige felszólító módú alakjával van dolgunk, azt egyértelműen mutatja a latin megfelelő: *Magister, aspice quales lapides et quales structurae.*

A pótkompetencia kiépítése Ágel 1988: 18 k. szerint legalább három síkon zajlik:

a/ *nyelven kívüli síkon*: ez kb. az adott szöveg korához tartozó történelmi-szociokulturális háttér tanulmányozását jelenti a pontosabb megértés érdekében;

b/ *metanyelvi síkon*: ez a vonatkozó szakirodalommal, valamint történeti nyelvtanokkal (pl. A magyar nyelv történeti nyelvtana) és szótárakkal (esetünkben pl. TESz., Károly 1965, Szabó T. 1977 stb.) való foglalkozáson keresztül vezet a minél pontosabb szövegértelmezés felé;

c/ *nyelvi síkon*: a vizsgált szövegen túl minél több közel abból az időből származó szöveget (pl. Bécsi Kódex, Apor-kódex, Jordánszky-kódex, Döbrentei-kódex stb.) kell tanulmányozni. Ez különösen ritka, esetleg a korpuszban csak egyszer előforduló, ma már érthetetlennek tűnő alakok esetén – megkönnyíti, ill. néha egyáltalában lehetővé teszi a megértést, sőt esetleg szövegek közötti egybevetésekre is módot ad.

Összegezve a fent elmondottakat, azt hiszem, megállapíthatjuk, hogy a kelőképpen kialakított pótkompetencia valóban viszonylag megbízható és objektív bázist tud nyújtani régi korok korpuszainak strukturalista indíttatású, pl. az igevalencia szemszögéből való leírásához.

IRODALOM

- Ágel, Vilmos 1988: *Überlegungen zur Theorie und Methode der historisch-synchronen Valenzsyntax und Valenzlexikographie: mit einem Valenzlexikon zu den „Denkwürdigkeiten der Helene Kottannerin (1439–1440)“*. *Lexicographica: Series Maior Bd. 25*, Tübingen, Niemeyer.
- Behrens, Leila 1989: *Textanalyse und Sprachwandel. Historische Syntax des Ungarischen: ein empirisches Modell. Europäische Hochschulschriften. Reihe XXI. Bd. 74*, Bern–Frankfurt am Main–Las Vegas, Lang.
- Closs, Elisabeth 1969: *Diachronic syntax and generative grammar*, in D. A. Reibel–S. A. Schane (szerk.): *Modern studies in English. Readings in transformational grammar*, Englewood Cliffs, Prentice-Hall.
- Coseriu, Eugenio 1975: *System, Norm und Rede*, in Coseriu: *Sprachtheorie und allgemeine Sprachwissenschaft*, München, Fink.
- Greule, Albrecht 1982: *Valenz, Satz und Text. Syntaktische Untersuchungen zum Evangelienbuch Otfrieds von Weißenburg auf der Grundlage des Codex Vindobonensis*, München, Fink.
- Greule, Albrecht szerk. 1982a: *Valenztheorie und historische Sprachwissenschaft. Beiträge zur sprachgeschichtlichen Beschreibung des Deutschen*, Tübingen, Niemeyer (Reihe Germanistische Linguistik, Band 42.)

- Károly Sándor 1965: *A Bécsi Kódex nyelvtana szótárszerű feldolgozásban*, Budapest, Akadémiai.
- King, Robert D. 1971: *Historische Linguistik und generative Grammatik. Übersetzt, eingeleitet und herausgegeben von Steffen Stelzer. Schwerpunkte Linguistik und Kommunikationswissenschaft. Bd. 5*, Frankfurt am Main.
- Korhonen, Jarmo 1978: *Studien zu Dependenz, Valenz und Satzmodell. Teil II. Untersuchung anhand eines Luther-Textes. Europäische Hochschulschriften. Bd. 271*, Bern–Frankfurt am Main–Las Vegas, Lang.
- Sitta, Horst 1971: *Semanteme und Relationen. Zur Systematik der Inhaltssatzgefüge im Deutschen*, Frankfurt.
- Szabó T. Ádám szerk. 1977: *Der Münchner Kodex IV. Wortschatz mit vollständigem Wort- und Formenverzeichnis*, Wiesbaden.
- Szemerényi Oszvald 1992: A modern nyelvtudomány fejlődésének a nyelvstruktúrában rejülő rugói, *Magyar Nyelv* 88, 1–13.
- TESz. = Benkő Loránd szerk. 1967–1976: *A magyar nyelv történeti-etimológiai szótára*, Budapest, Akadémiai.
- Van de Velde, Roger G. 1971: *Zur Grundlegung einer linguistischen Methodik. Gezeigt am Beispiel der altfriesischen Syntax. Commentationes Societatis Linguisticae Europae IV*, München.

GESCHLOSSENE CORPORA UND ERSATZKOMPETENZ TAMÁS FORGÁCS

In der neueren Sprachwissenschaft gibt es immer mehr Versuche, Methoden der – meist auf die Synchronie gerichteten – modernen Forschungsrichtungen (strukturalistische und generative Grammatiken) auch diachron einzusetzen. So gibt es Versuche für die diachrone Anwendung des generativen Modells, noch mehr kommt aber die Valenztheorie für die Erforschung der grundlegenden Satzstrukturen bzw. Satzmodelle historischer Sprachstufen zum Einsatz.

Wegen der Exaktheit solcher Untersuchungen muß aber die Problematik gründlich unter die Lupe genommen werden, wie und in welchem Maße gewisse Handicaps der diachronen Analyse gegenüber der synchronen bewältigt werden können. Zum einen ist hier die Frage der sog. geschlossenen Corpora, zu denen keine Informanten zu erreichen sind, zu untersuchen, zum anderen – in unmittelbarem Zusammenhang damit – die Frage der Kompetenz, die in diesem Falle nur eine – vom Deskribenten aufgebaute – Ersatzkompetenz sein kann.

Der Autor faßt in seiner Arbeit diese Problematik zusammen und stellt auch die Lösungsvorschläge vor, die in der Literatur zu finden sind.