

# Az ellenforradalmár

Kornai András

Algebra Tanszék, Budapesti Műszaki és Gazdaságtudományi Egyetem

1111 Budapest Egry J. u. 1

kornai@math.bme.hu

## Kivonat

Cikkünkben rendszeres áttekintést nyújtunk Chomskynak a nyelv-elmélet formalizálására irányuló korai munkáiról. Utólag könnyű látni, hogy paradigmateremtő munkássága nemcsak egyoldalú volt, amennyiben a nyelvnek és a nyelvtudománynak mindenestül a diszkrét aspektusaira koncentrált, de aktívan ellenséges is a folytonos aspektussal, különösen a gyakorisági adatok használatával szemben.

A statisztikai gondolkodás Maxwell, Boltzmann és Gibbs termodinamikai munkájával a XIX. század második felében jelent meg a fizikában, és a kvantumelmélet megjelenésével elkezdte áthatni annak minden aspektusát. Ugyanebben az időszakban statisztikai alapokra került a biztosítási matematika, az evolúcióelmélet, a járványtan (és általában az orvostudomány), az ökonometria, a pszichológia és a társadalomtudomány nagy része is. A statisztikai és valószínűségszámítási megfontolások olyan mértékben átjárták a tudományos gondolkodást, hogy

röviddel 1930 után gyakorlatilag biztossá vált, hogy világunk alapjait legjobb esetben is a véletlen törvényei működtetik. Az episztemológiában [...] ma közhely, hogy a tapasztalatból való tanulásunknak, sőt a tudásunk alapjainak a nagy részét is valószínűségi modellekkel kell ábrázolni. (Hacking, 1987)

Ezzel az általános háttérrel szemben roppant figyelemre méltó, hogy az 1950-es és 1960-as években egy ragyogó, sok szempontból radikális és forradalmi gondolkodó, a fiatal Noam Chomsky, megpróbált ennek egymaga gátat

vetni, és a nyelvészetet a determinisztikus elméletalkotás utolsó végvárává tenni. Ebben jelentős sikert aratott, hiszen nyelvészek nemzedékei nevelkedtek föl elfogadva a valószínűségi modellek használatával szembeni érveit. Itt kevésbé foglalkozunk azzal, hogy miért tette ezt (voltak a valószínűségnek más nevezetes ellenzői is, mint Einstein, aki 1955-ben bekövetkezett haláláig vívta vesztes csatáját a kvantummechanikával szemben), inkább Chomsky érvanyagára összpontosítunk, amelyet figyelemre méltó cikkekben és könyvekben sorakoztatott föl az 1950-es évek végén és az 1960-as évek elején.

A kezdetek már jelen vannak PhD-disszertációjában (1954-1955), amit Chomsky, (1975)-ként közöltek rövidített formában, és néhány belső jelentésben, amelyek az MIT Elektronikai Kutatólaboratóriumának *Quarterly Progress Report*ájában jelentek meg (1956, 41. és 42. szám). Nem látjuk okát, hogy ezeket részletesen tárgyaljuk, mivel Chomsky elsőbbségét ezekben a kérdésekben senki sem vitatja, és csak az ezt követő, szélesebb körben hozzáférhető, a kezdeti érvelést megismétlő és kiterjesztő publikációk gyakoroltak messzemenő hatást. (A Google Tudós szerint Chomsky első, a témába vágó folyóirat-publikációját (Chomsky, 1956) 3212-szer idézték, míg a belső jelentéseket összesen 12-szer.)

Amint látni fogjuk, a legtöbb érve a végtelennek vélelmezett sorozatok természetével és bizonyító erejével kapcsolatos egyszerű, de gyakori félreértéseken alapul, míg a többiek olyan formális eszközöket támadnak, melyek abban az időben még lényeges szempontokból éretlenek voltak.

## 1. Az információelmélet elleni küzdelem

A kommunikáció valószínűségi elmélete Shannonnal (1948) kezdődik. Állítható, hogy a híres  $H = -\sum_i p_i \log_2 p_i$  entrópiaképlet még az  $E = mc^2$ -nél is szerveesebb része annak a képlettárnak, amelyen a 21. századi szellemi fejlődés alapszik, hiszen a bitek és a bájtok ugyanannyira a mindennapi élet részévé váltak, mint a voltok és az amperek, míg a radok és a grayek hálaistennek nem. Figyelemreméltó, hogy az elmélet matematikai alapjai hozzáférhetőek minden gimnazista számára akik már jól értik a logaritmust, és az 1950-es években általánosan tanították az MIT-n. Ami azt illeti, a téma vezető helyi tekintélye, Robert Fano, Shannon egyik társszerzője – akinek a tankönyvét (Fano, 1961) széles körben használták, amíg Cover and Thomas, (1991) fel nem váltotta – történetesen ugyanabban a 20-as épületben dolgozott évtizedeken át, mint Chomsky. Mindent összevéve Chomskynak egyértelműen

megvoltak az eszközei és a lehetősége, hogy teljes mértékben megértse a témát, semmiképp nem fogható vak tudatlanságra az, hogy totális dezinformációs kampányt folytatott az információelmélet ellen.

Egy tipikus példával kezdjük Chomskynak Greenberg, (1957)-ről írt recenziójából. A 2. lábjegyzetben Chomsky, (1959) így ír:

A feltételezés az, hogy a mondat minden pontján van egy bizonyos számú fennmaradó lehetőség, hogy mi lesz a mondat; és hogy ha egy nagyobb osztályból választunk egy elemet, az jobban csökkenti ezt a számot, mint egy kisebb halmazból való választás, és így több „információ”-t ad a hallgatónak arról, hogy melyik lesz a szóban forgó mondat. Ez az értelmezés azonban általában elfogadhatatlan. Mivel tipikusan egy mondatban egy adott ponton végtelen számú lehetséges sorozat van, amellyel befejeződhet a mondat, egy adott elem választása egyáltalán nem csökkenti a fennmaradó alternatívák számát.

Egyértelmű, hogy Greenbergnek igaza volt, Chomsky pedig tévedett, mivel nem a pusztán számra, hanem az információra megy ki a játék. Egy olyan mondat, amely így kezdődik: *Hé, hallottad, hogy* és egy olyan, ami így: *Sokféleképpen lehet egyszerűen*, egy-egy valószínűségi eloszlást generál a lehetséges következő szavak halmazán, és az előbbi entrópiája 4.6 bit, ami lényegesen nagyobb, mint az utóbbié, ami 2.9 bit. Egy egyszerű, statisztikai nyelvmodellt használva, amelyet a Penn Treebank-en (Marcus, Santorini, and Marcinkiewicz, 1993) tanítottunk fel, az első szósor angol változata után a legvalószínűbbnek jósolt szó a *the* ( $p = 0,27$ ), az utóbbié után pedig a *to* ( $p = 0,35$ ), míg a *to* valószínűsége az első esetben 0.006 és a *the*-é a másodikban 0.012. Ezen a ponton még nehéz megmondani, mi vezeti Chomskyt: szándékos tudatlanság (ezt nehéz elképzelni) vagy tudatos választás, de az idő múltán az érvek explicitebbek lettek, ezért az utóbbit fogjuk feltenni. A sokat idézett Chomsky, (1956) nyíltan szembeszáll az egyszerű *n-gram* modellel (2.4. szakasz):

Feltehetjük a kérdést, hogy lehet-e konstruálni [végesállapotú nyelvtanokból] egy olyan sorozatot, amely valamilyen nem-triviális módon egyre jobban közelíti egy kielégítő angol nyelvtan kimenetét. Tegyük fel például, hogy rögzített  $n$  esetén a következőképpen készítünk véges állapotú nyelvtant: minden angol szavakból álló

$n$ -hosszú sorozatnak megfelel a nyelvtan egy állapota, és  $X$  szó kibocsátásának a valószínűsége, amikor a rendszer az  $S_i$  állapotban van, egyenlő az  $X$  feltételes valószínűségével az  $S_i$ -t definiáló  $n$  szószorozat mint feltétel mellett. Az ilyen nyelvtan kimenetét általában az angol  $n + 1$ -ed rendű közelítésének hívják. Ahogy  $n$  növekszik, az ilyen nyelvtanok kimenete nyilvánvalóan egyre jobban úgy fog kinézni, mint az angol, mivel egyre hosszabb szekvenciákat fogunk nagy valószínűséggel közvetlenül az angolnak abból a mintájából venni, amelyben a valószínűségeket megállapítottuk. Ez a tény alkalmanként ahhoz a javaslatához vezetett, hogy a nyelvi struktúra elméletét ilyen modellek szerint kellene kialakítani.

Bármilyen egyéb érdeklődés fűződjön is az ilyen jellegű statisztikai közelítéshez, egyértelmű, hogy nem világíthat rá nyelvtani problémákra. Nincs általános összefüggés egy sor (vagy alkotóelemeinek) gyakorisága és grammatikussága között. Ezt a választóvonalat egyértelműen láthatjuk, ha olyan sorokat veszünk példaként, mint *színtelen zöld ötletek alszának dühösen* ami egy grammatikus mondat, noha okkal feltételezhetjük, hogy a benne szereplő szavak közül egyetlen pár sem jelenhetett meg soha együtt a múltban ...

Azokban az időkben kevés nyelvész rendelkezett megfelelő statisztikai háttérrel, hogy kiszúrja a hibát, nevezetesen, hogy egy karaktorsor valószínűsége nem egyenlő a korpuszban való gyakoriságával, bármilyen nagy legyen is az, hiszen a korpusz csak véges minta egy végtelen eloszlásból. Míg a *színtelen zöld ötletek* érvelését célzó részletes kritikának várnia kellett Pereira, (2000)-ig, a beszédtechnológusok (Jelinek and Mercer, 1980) és a karakterfelismeréskutatók közössége (Kornai, 1994) megőrizte „egyéb érdeklődését” a következő szó valószínűségének becslése iránt: ezt a feladatot általában statisztikai *nyelvmodellezésnek* hívjuk. Ezekben a közösségekben és az információkinyerés (Ponte and Croft, 1998) és a gépi fordítás (Brown et al., 1990) világában az elejétől nyilvánvaló volt, hogy zérus becslés nem elfogadható akkor sem, ha zérus gyakoriságot látunk, az eloszlást *simítani* kell. Történelmi érdekesség gyanánt megemlítjük, hogy az első széles körű simító módszer nagyrészt Fred Jelineknek köszönhető, akinek Robert Fano volt a doktori témavezetője.

Nézzük meg most Chomsky and G. Miller, (1958)-at, az *Information and Control* első évfolyamának egyik nagyon vonzó cikkét, amely egyértelműen

zászlóshajóként segítette útjára az  $I\mathcal{E}C$ -t mint a formális nyelvelmélet legkiválóbb folyóiratát. A cikk többek között tartalmazza a véges állapotú nyelvek (*finite state language*, FSL) első szisztematikus vizsgálatát, több olyan tételt bizonyít, amelyek máig a formális nyelvészeti tananyag szabvány részei, pl. hogy az FSL-ek családja zárt a Boole-műveletekre.

A véges állapotú automaták (*finite state automaton*, FSA) akkoriban nagy figyelmet kaptak, McCulloch and Pitts, (1943) agymodelljének FSA-ként való meghatározó átfogalmazása (Kleene, 1956) tisztelethelyet kapott a nagy hatású C. Shannon and McCarthy, (1956) tanulmánykötetben, ahol egyébként az egyenértékű Moore-gépeket (Moore, 1956) bevezették. A Mealy-gépeket, egy harmadik egyenértékű megfogalmazást, az előző évben vezette be Mealy, (1955). Ezek mind determinisztikus modellek voltak: a legfontosabb megfigyelést, hogy a nem-determinizmus (szabad akarat) nem változtatja meg az elfogadott/generált nyelvek osztályát, végül Rabin and Scott, (1959) bizonyította.

Teljesen egyértelmű, hogyan illeszkedik Chomsky and G. Miller, (1958) a Chomsky-féle dezinformációs kampány vonalába:

Az utóbbi években ismerős eljárásá vált az üzenetforrások sztochasztikus folyamatként – rendszerint véges Markov-folyamatként – való ábrázolása (C. E. Shannon, 1948). A véges Markov-folyamat lényegében egy véges állapotú generátor, kiegészítve minden állapotban a rendelkezésre álló lehetőségek fölötti valószínűségi eloszlással. Két fontos kérdés merülhet fel az ilyen generátorokkal kapcsolatban: (1) Milyen tulajdonságok jellemzik az általuk előállított nyelveket, és (2) A természetes nyelvek rendelkeznek-e ezekkel a tulajdonságokkal? Úgy tűnik, hogy a véges állapotú nyelvten nem alkalmas a legtöbb természetes nyelv leírására (Chomsky, 1956), tehát a második kérdésre a válasz nemleges. Mindazonáltal az ilyen eljárások matematikai tulajdonságai jól megfelelnek a hírközlési mérnökök igényeihez, és valószínűleg továbbra is érdekesek lesznek a kommunikációelmélet számos alkalmazásában.

Mivel az akkori vezető agymodell az FSA-val egyenértékű volt, a logikus következtetés az volt, hogy a természetes nyelvek véges állapotúak. Chomsky, (1957) nagy erőfeszítéseket tett, hogy ezt – a középponti beágyazásos konstrukciók segítségével – cáfolja, és később további epiciklusokkal egészí-

tette ki az elméletet. Ezek közül a legfontosabb, hogy az egyébként egyértelműen szükséges kompetencia/performancia megkülönböztetést (Chomsky, 1965) kiterjesztette a középponti beágyazásokra csak azért, hogy megvédje ezt a bizonyítását. Az érvényességi területnek ez a kiterjesztése a kezdetektől meglehetősen megosztónak bizonyult (és nehéz bárkit is találni a Chomsky-hívek egyre szűkülő csoportján kívül, aki továbbra is elfogadja), de ezt a jól dokumentált vitát itt félretesszük, és teljes mértékben a valószínűségi szempontokra összpontosítunk.

Nem más ez, mint egy arra irányuló módszeres erőfeszítés, hogy a Markov-rendszereket olyan FSA-ként mutassa be, amelyeket a valószínűségi eloszlások „kiegészítenek”, majd a valószínűségi aspektus nélkül tanulmányozzuk az FSA-kat, hisz ez az aspektus Chomsky szerint kiegészítő jellegű. Chomsky and G. Miller, (1958) utolsó szakaszai, ahol a  $\lambda$ -hosszúságú mondatokat számolja meg, már ezt a stratégiát mutatja: a gyakoriságokat egy ügyes bővésztrükkel eltünteti. Először bizonyos mondatokat önkényesen explicit végpontokkal ellátott mondatokkal helyettesít. Ez megtehető az általánosság megszorítása nélkül az 1. tétel szerint, amely csak a gyakoriságoktól megfosztott sztringhalmazokkal foglalkozik. Ezután a véges állapotú nyelvtanokat helyettesíti egyértelműekkel, ismét az általánosság megszorítása nélkül a 3. tétel szerint mindaddig, amíg a határjelzők jelen vannak. Miután a szemfényvesztés befejeződött, C. E. Shannon, (1948) módszerével megszámlálja a  $\lambda$ -hosszú sztringeket. Az egész itt-a-piros elfogadhatatlan, hiszen léteznek olyan valószínűségi nyelvek véges állapotú tartóval, amiket nem lehet „valószínűségekkel kiegészített” FSG-kkel generálni, ám ezt csak egy évtizeddel később bizonyította Ellis, (1969). Figyelemre méltó utóvédharcot végzett Suppes, (1970), Levelt, (1974) pedig ezt kanonizálta azzal, hogy egyszerűen tételként közölte Suppes hamis sejtését – a részleteket lásd Kornai, (2011)-ben.

## 2. Mi történt?

Amit itt dokumentálunk, nem pusztán hiba. Aki ismeri Chomskynak a formális nyelvtanokra vonatkozó munkásságát (olyan területről van szó, mely alapjában tőle ered), egyszerűen nem hitetheti el magával, hogy Chomsky valahogy mellégondolkodta magát, szellemileg megbotlott, mint ahogy el-írunk egy szót. Utólag nyilvánvaló, hogy Chomsky karnyújtásnyira ment el a 20. század néhány legnagyobb felfedezése mellett, a non-determinizmus

(Rabin and Scott, 1959), és a rejtett Markov-modellek (HMM, Stratonovich, 1960; Baum and Petrie, 1966) mellett. Ő is ugyanezekkel a kérdésekkel foglalkozott, és a valószínűségi modelleknek a G. A. Miller and Chomsky, (1963)-ben bemutatott kiterjedt kezelése (70 oldal, kismonográfiai terjedelem) bőséges bizonyíték matematikai szakképzettségére. Ám igyekezetében, hogy a saját transzformációs nyelvtanfoglalmát úgy állítsa be, mint a természetes nyelv egyetlen alkalmas modelljét (amely kísérletre Peters and Ritchie, (1973) jóval a valószínűségi modellek újjászületése előtt végső csapást mért), kihagyta nemcsak a non-determinizmust, amely rendkívül fontos általában a mesterséges intelligenciához és különösen a programnyelvek szemantikájához (Floyd, 1967), hanem – amint alább igyekszünk bemutatni – a HMM-eket is, amiket pedig messzemenően használnak nemcsak a természetes nyelv tanulmányozásánál, hanem a természet nyelvénél, a genetikai kódnál is (Durbin et al., 1998).

Nehéz mások gondolkodását kitalálni, de nincs jó okunk feltételezni, hogy a dezinformációs kampányt Chomsky rosszhiszeműen folytatta. Valószínűbb, hogy alaposan átgondolta ezeket a dolgokat, és először is magát győzte meg arról, hogy a valószínűségek irrelevánsak a nyelvtanban. Bepillantást enged a gondolatmenetébe Chomsky, (1956) már idézett 2.4. szakasza: „a »színtelen zöld ötletek alszanak dühösen« (...) grammatikus (...) noha okkal feltételezhetjük, hogy *a benne szereplő szavak közül egyetlen pár sem jelenhetett meg soha együtt a múltban*” (kiemelés tőlem). Nyilvánvaló, hogy a „színtelen zöld”, a „zöld ötletek”, az „ötletek alszanak” (*ideas sleep*) és az „alszanak dühösen” (*sleep furiously*) bigramok mind rendkívül ritkák.

Ma, hogy már igazán nagy méretű korpuszok állnak rendelkezésünkre (az alábbi keresési eredmények a Google Könyvekből vannak), találunk egy példát a „színtelen zöld”-re a *School of Mines Quarterly* bányászati folyóirat 8. kötetében (1887) egy minőségi fúvócsöves elemzésről szóló cikkben (363. o.); és a „Remek zöld ötletek: Környezeti nevelés gyerekeknek” című felülírja azt, ami elsőre nagyon erős kiválasztási korlátozásnak tűnik, amely kizárná a színnevek használatát elvont tárgyakkal. Miközben a XX. századi anyagot Chomsky példája uralja, egy 1868-as könyvben ezt találjuk: „pihent abban a titokzatos hálóteremben, ahol az ötletek alszanak (*ideas sleep*), készen arra, hogy életre ébredjenek”, egy 1932-es kötetben pedig ezt: „Akiknek egy napot az árkokban kell tölteniük, utána dühösen alszanak (*sleep furiously*)”. Minden olyan generatív nyelvész számára, aki az intuícióira támaszkodik, ismerős ez a helyzet: esküszünk, hogy valami nem grammatikus, amíg nem szembesülünk a valódi életből vett példákkal. Jó bízni az intuíciónkban,

de még jobb verifikálni.

Ezekből a példákból nem csak azt a trivialitást vihetjük magunkkal, hogy a „nem látott” (*unattested*) nem helyettesíti a „nem grammatikus”-t, hanem azt is megmutatják, hogy Chomsky egy paraszthajszálnyira járt a Rejtett Markov áttöréstől. Tudta, hogy a melléknév-melléknév, melléknév-főnév, főnév-ige, ige-határozó osztálybigramok teljesen szokásosak, és az a kritikus intuíciója is megvolt, hogy ez a mondat igenis grammatikus. Mindössze annyi hiányzott, hogy a mély és felszíni struktúra megkülönböztetéséről való saját elképzelését átvigye a preterminálisok (szófaji kategóriák) és szótári bejegyzések területére!

90. születésnapja alkalmából talán illőbb lenne Chomsky maradandó eredményeit ünnepelni, mint pártatlanul, *sine ira et studio* elemezni egy olyan kudarcot, melyet nála kisebb elmék könnyebben elkövettek volna. De a dicsérő irodalom hatalmas, és szinte mindig ugyanazokat a hibákat követi el, mint a mester, ez pedig felveti a kollektív felelősség kérdését. Chomsky elmulasztott karnyújtásnyira levő nagy felfedezéseket, s ezzel már drágán megfizetett az ellenforradalmi álláspontjáért, a vád további része a vak követőket terheli. Ahelyett, hogy a 21. századi kutatás létfontosságú részévé vált volna, a kilencvenes évekre a klasszikus generatív mondattan egyre inkább a rozsdáövezet városaihoz hasonlított, ahol a helyi ipar halott vagy lassan haldoklik, a régi szállítógányok használaton kívül kerültek, és nincs remény a város felélesztésére. A forradalmat nem lehetett megállítani.

## Köszönetnyilvánítás

Köszönöm Paul Kiparskynak, hogy lényegbevágó megjegyzéseket fűzött egy korábbi változathoz. A fordító háláját fejezi ki Nádasdy Ádámnak, aki részletes megjegyzésekkel segítette a szöveg folyékonyságát és helyességét, beleértve az ominózus zöld ötleteket.

## References

Baum, Leonard E. and Ted Petrie (1966). “Statistical inference for probabilistic functions of finite state Markov chains”. In: *Annals of Mathematical Statistics* 37, pp. 1554–1563.



- Brown, Peter et al. (1990). “A statistical approach to machine translation”. In: *Computational Linguistics* 16, pp. 79–85.
- Chomsky, Noam (1956). “Three models for the description of language”. In: *IRE Transactions on Information Theory* 2, pp. 113–124.
- (1957). *Syntactic Structures*. The Hague: Mouton.
- (1959). “On certain formal properties of grammars”. In: *Information and Control* 2, pp. 137–167.
- (1965). *Aspects of the Theory of Syntax*. MIT Press.
- (1975). *The logical structure of linguistic theory*. Springer Verlag.
- Chomsky, Noam and George Miller (1958). “Finite-state languages”. In: *Information and Control* 1, pp. 91–112.
- Cover, Thomas M. and Joy A. Thomas (1991). *Elements of Information Theory*. Wiley.
- Durbin, R. et al. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press.
- Ellis, Clarence A. (1969). *Probabilistic Languages and Automata*. University of Illinois, Urbana: PhD thesis.
- Fano, Robert (1961). *Transmission of Information: A Statistical Theory of Communications*. MIT Press.
- Floyd, Robert W (1967). “Nondeterministic algorithms”. In: *Journal of the ACM (JACM)* 14.4, pp. 636–644.
- Greenberg, Joseph H., ed. (1957). *Essays in linguistics*. University of Chicago Press.
- Hacking, Ian (1987). “Was There a Probabilistic Revolution 1800-1930?” In: *The Probabilistic Revolution*. Ed. by L. Krüger et al. Cambridge MA: MIT Press.
- Jelinek, Frederick and Robert Mercer (1980). “Interpolated estimation of Markov source parameters from sparse data”. In: *Proceedings of the Workshop on Pattern Recognition in Practice*. Ed. by E. S. Geltsema and L. N. Kanal. Amsterdam: North-Holland.
- Kleene, Stephen C. (1956). “Representation of events in nerve nets and finite automata”. In: *Automata Studies*. Ed. by C. Shannon and J. McCarthy. Princeton University Press, pp. 3–41.
- Kornai, András (1994). “Language models: where are the bottlenecks?” In: *AISB Quarterly* 88, pp. 36–40.
- (2011). “Probabilistic grammars and languages”. In: *Journal of Logic, Language, and Information* 20, pp. 317–328.

- Levelt, Willem J.M. (1974). *Formal Grammars in Linguistics and Psycholinguistics*. Vol. 1–3. The Hague: Mouton.
- Marcus, Mitchell, Beatrice Santorini, and Mary Ann Marcinkiewicz (1993). “Building a Large Annotated Corpus of English: The Penn Treebank”. In: *Computational Linguistics* 19, pp. 313–330.
- McCulloch, W.S. and W. Pitts (1943). “A logical calculus of the ideas immanent in nervous activity”. In: *Bulletin of mathematical biophysics* 5, pp. 115–133.
- Mealy, George H. (1955). “A method for synthesizing sequential circuits”. In: *Bell System Technical Journal* 34, pp. 1045–1079.
- Miller, George A. and Noam Chomsky (1963). “Finitary models of language users”. In: *Handbook of Mathematical Psychology*. Ed. by R.D. Luce, R.R. Bush, and E. Galanter. Wiley, pp. 419–491.
- Moore, E.F. (1956). “Gedanken-experiments on sequential machines”. In: *Automata studies*. Ed. by Shannon and McCarthy. Princeton University Press, pp. 129–153.
- Pereira, Fernando (2000). “Formal grammar and information theory: Together again?” In: *Philosophical Transactions of the Royal Society A* 358, pp. 1239–1253.
- Peters, Stanley and Robert W. Ritchie (1973). “On the generative power of transformational grammars”. In: *Information Sciences* 6, pp. 49–83.
- Ponte, Jay M. and W. Bruce Croft (1998). “A language modeling approach to information retrieval”. In: *Proc SIGIR*. ACM Press, pp. 275–281.
- Rabin, M.O. and D. Scott (1959). “Finite automata and their decision problems”. In: *IBM journal of research and development* 3.2, pp. 114–125. ISSN: 0018-8646.
- Shannon, Claude E. (1948). “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27, pp. 379–423, 623–656.
- Shannon, Claude and John McCarthy, eds. (1956). *Automata studies*. Princeton University Press.
- Stratonovich, R. L. (1960). “Conditional Markov Processes”. In: *Theory of Probability and Its Applications* 5.2, pp. 156–178.
- Suppes, Patrick (1970). “Probabilistic grammars for natural languages”. In: *Synthese* 22, pp. 95–116.