Harris, Firth, and Distributional Semantics

Dirk Geeraerts

PREVIEW

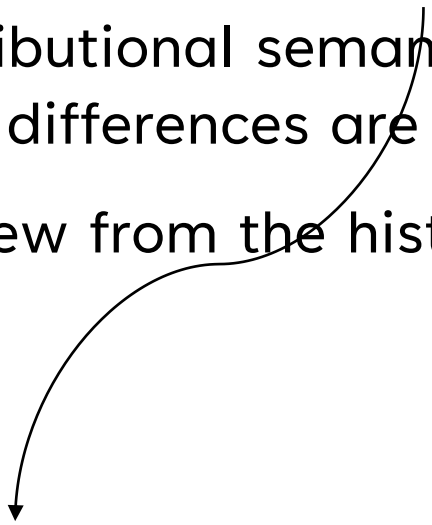
Harris and Firth are often mentioned as founding figures of distributional semantics, but their actual impact and mutual differences are seldom examined

→ a view from the history of linguistics

PREVIEW

Harris and Firth are often mentioned as founding figures of distributional semantics, but their actual impact and mutual differences are seldom examined

→ a view from the history of linguistics

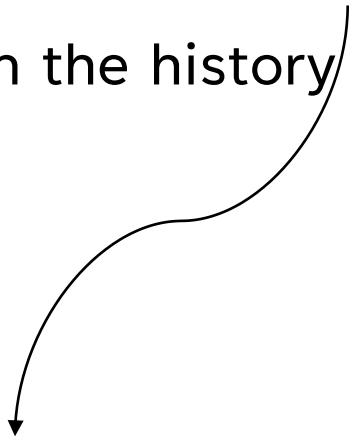


rough count:
25% of ‘distributional semantics’
publications of past 5 years

PREVIEW

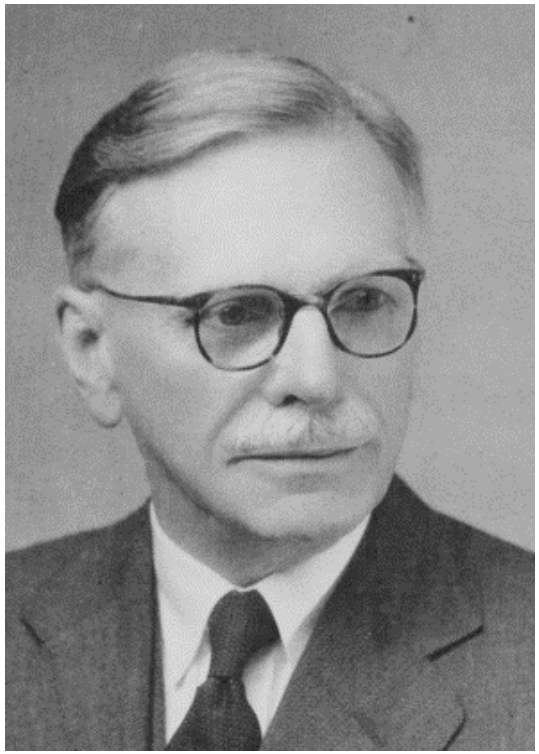
Harris and Firth are often mentioned as founding figures of distributional semantics, but their actual impact and mutual differences are seldom examined

→ a view from the history of linguistics



but see (from a computational perspective) a.o.
- Brunila & LaViolette 2022, 'What company do words keep?'. arXiv:2205.07750
- Léon 2021, *Automating Linguistics*. Springer.
- Lenci & Sahlgren 2023. *Distributional Semantics*. CUP.

John Rupert Firth
1890-1960



Zellig Harris
1909-1992

main sources:

- Firth 1935 ‘The technique of semantics’
  Firth 1950 ‘Personality and language in society’
  Firth 1957 ‘A synopsis of linguistic theory 1930-1955’
- Harris 1951, *Structural Linguistics*
  Harris 1954, ‘Distributional structure’
  (i.e. only the early Harris)

PREVIEW

steps

- distributional semantics is more than vector space semantics
- neither Harris nor Firth are direct founding figures of vector space semantics
- indirectly, both have been influential for distributional semantics at large (but Firth more so than Harris)
- tracing these lineages helps identify issues for vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

describing the meaning of an expression on the basis of the elements it occurs with involves two steps:

- selecting and identifying relevant contextual features of targets
  (primary distribution)
- finding distributions of distributions, i.e. patterns in the co-occurrences
  (secondary distribution)

both steps can be taken manually or statistically

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*		
	statistical		vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*		
	statistical		vector space semantics

	predict	count
static		
contextual		

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

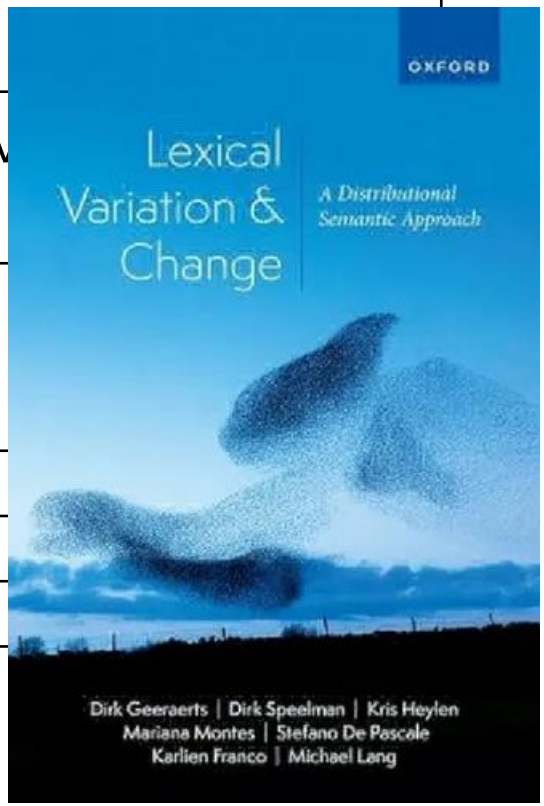
		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*		
	statistical		vector space semantics

	predict	count
static	Word2Vec, GloVe	LSA, HAL
contextual	ElMo, BERT, GPT	

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*		
	statistical		

	predict	count
static	Word2Vec, GloVe	LSA, HAL
contextual	ElMo, BERT, GPT	



1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*		
	statistical		vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	
	statistical		vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical		vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical		vector space semantics

1990s corpus lexicology:
corpus-driven description of lexical co-occurrence patterns
(collocation, colligation, semantic prosody)
e.g. KWIC concordance + manual interpretation

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical		vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

manual or semi-automatic annotation of usage cases
for a variety of features + multifactorial analysis
emerged in the context of cognitive linguistics;
now standard in usage-based linguistics and construction grammar
Gries 2000, Grondelaers 2000, Divjak 2006, Glynn 2008 (etc.)

1 VARIETIES OF DISTRIBUTIONAL SEMANTICS

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

+ intermediate and hybrid approaches

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

both H and F provide excellent inspirational aphorisms

- Firth 1957
  ‘You shall know a word by the company it keeps’
- Harris 1954
  ‘Difference of meaning correlates with difference of distribution’

but …

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

1) contextual meaning determination was not invented by H or F; cp. the lexicographical and philological tradition

   neither was distributional analysis; cp. the legacy of structuralist phonology

   minimal pairs: top / tip
   complementary distribution: top / pot

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

2) neither H nor F developed an actual statistical corpus analysis

(even though H, inspired by Shannon, explicitly thinks in probabilistic terms)

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

3) neither H nor F is primarily interested in methods of lexical analysis

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

3) neither H nor F is primarily interested in methods of lexical analysis; re **Harris**:

purpose is identifying the formal structure of a language with a procedure that is
1° independent of (semantic) intuition

> "here we will discuss how each language can be described in terms of a distributional structure (...) and how this description is complete without intrusion of other features such as history or meaning" 1954: 146

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

3) neither H nor F is primarily interested in methods of lexical analysis; re **Harris**:

purpose is identifying the formal structure of a language with a procedure that is
2° multi-tiered and incremental

> "when the distribution of phonemes is considered over longer stretches of speech they are found to be highly restrictive; (...) we can state these limitations by setting up new (morphemic) elements"
> 1951: 156

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

3) neither H nor F is primarily interested in methods of lexical analysis; re **Harris**:

purpose is identifying the formal structure of a language with a procedure that is
2° multi-tiered and incremental

- the distribution of sounds identifies phonemes
- the distribution of phonemes identifies morphemes
- meaning only plays a minimal role at the very start of the procedure

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

3) neither H nor F is primarily interested in methods of lexical analysis; re **Firth**:

purpose is description of language in contexts of situation (cp. Malinowski, Wittgenstein)

> "the central concept of semantics considered this way is the context of situation"
> 1957(1935): 27

> "a bit of the social process which can be considered apart and in which a speech event is central"
> 1957(1950): 182

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

3) neither H nor F is primarily interested in methods of lexical analysis; re **Firth**:

purpose is description of language in contexts of situation with collocational behavior as indicative

> "The day-to-day practice of playing language games recognizes customs and rules. It follows that a text in such established usage may contain sentences such as *Don't be such an ass!, You silly ass!, What an ass he is!* In these examples, the word *ass* is in familiar and habitual company, commonly collocated with *you silly—, he is a silly—, don't be such an—*" 1962(1957): 11

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

4) H and F disagree in fundamental respects; viz. F explicitly opposes H's meaning-free, bottom-up procedure

> "This does not mean that the analysis of discourse (...) can be directly developed from phonemic procedures or even devised by analogy from such procedures"
>
> [in reference to H 1951]
>
> 1962(1957): 21

2 FOUNDING FIGURES OF VECTOR SPACE SEMANTICS?

shallowly referring to H and F as ‘founding fathers’ paints a somewhat distorted picture

- of their specific interests and outlook
- of the fundamental differences between both
- of their direct influence on vector space semantics
- of its prehistory and wider context

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

Firth’s influence can be traced along two trajectories, a linguistic and a computational one

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

a linguistic trajectory:

Sinclair 1987
Collins Cobuild Dictionary of Contemporary English

corpus-based
focus on collocations
- lexical usage patterns as descriptive target
- sentential explanations instead of
dictionary-style definitions

e.g. **contact**
2 If there is eye contact between people, they are looking straight at one another

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

FIRTH

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

FIRTH → SINCLAIR (COBUILD)

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

a linguistic trajectory:

1990s
‘neo-Firthian corpus linguistics’
(a.k.a. the Birmingham school of corpus linguistics)

John Sinclair,
Michael Stubbs, Susan Hunston, Michael Hoey,
Wolfgang Teubert, Elena Tognini-Bonelli

corpus-driven description of co-occurrence
patterns (collocation, colligation, semantic prosody)

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

FIRTH → SINCLAIR (COBUILD)

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

FIRTH → SINCLAIR (COBUILD)

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

a linguistic trajectory:

2000s
incorporation of corpus methods in cognitive linguistics
(a.k.a. ‘the quantitative turn in cognitive linguistics’)

expansion of corpus methodology

- statistical techniques (logistic / mixed effects regression, correspondence analysis ...)

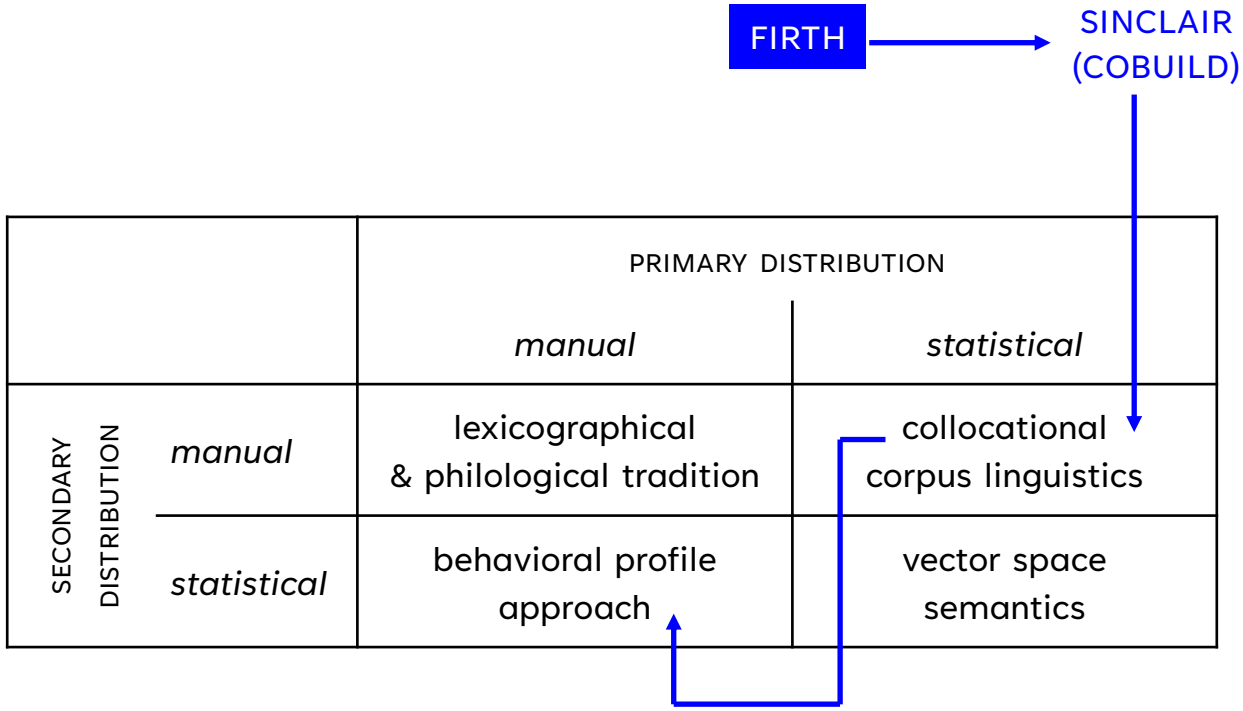
- range of collocational phenomena (behavioral profiles, collostructions)

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

FIRTH → SINCLAIR (COBUILD)

		PRIMARY DISTRIBUTION	
		manual	*statistical*
SECONDARY DISTRIBUTION	*manual*	lexicographical & philological tradition	collocational corpus linguistics
	statistical	behavioral profile approach	vector space semantics

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH



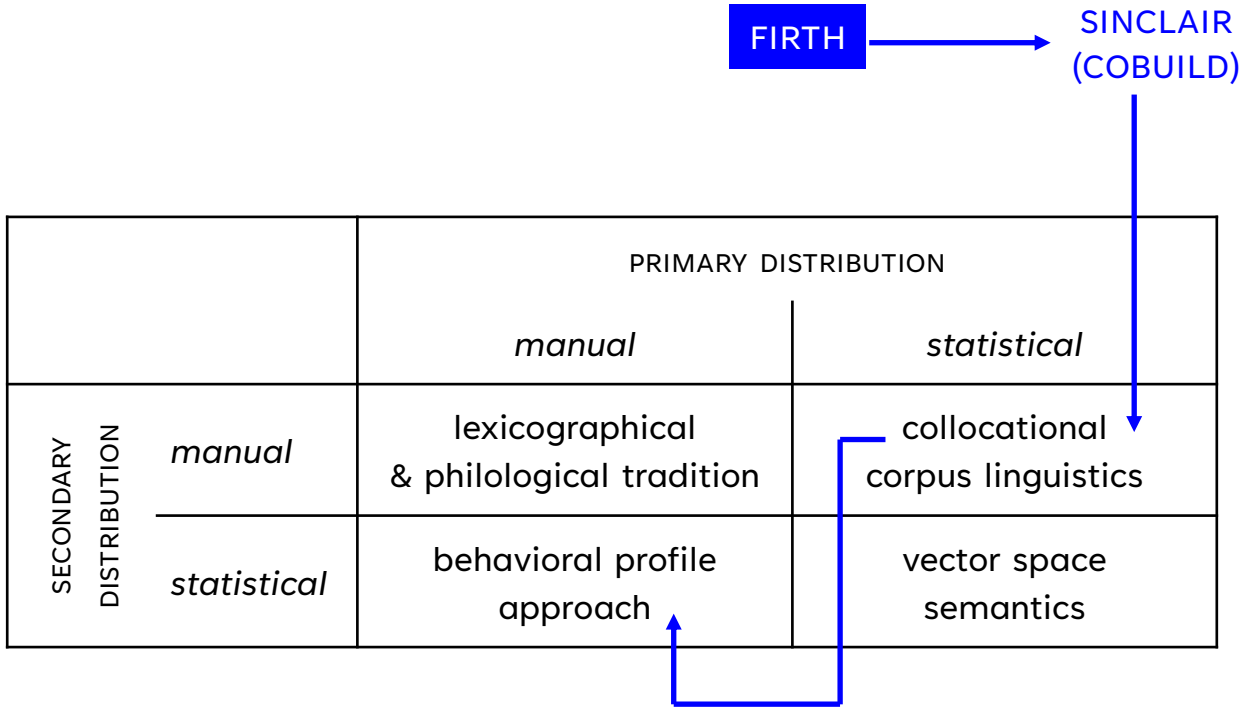
3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

a computational trajectory:

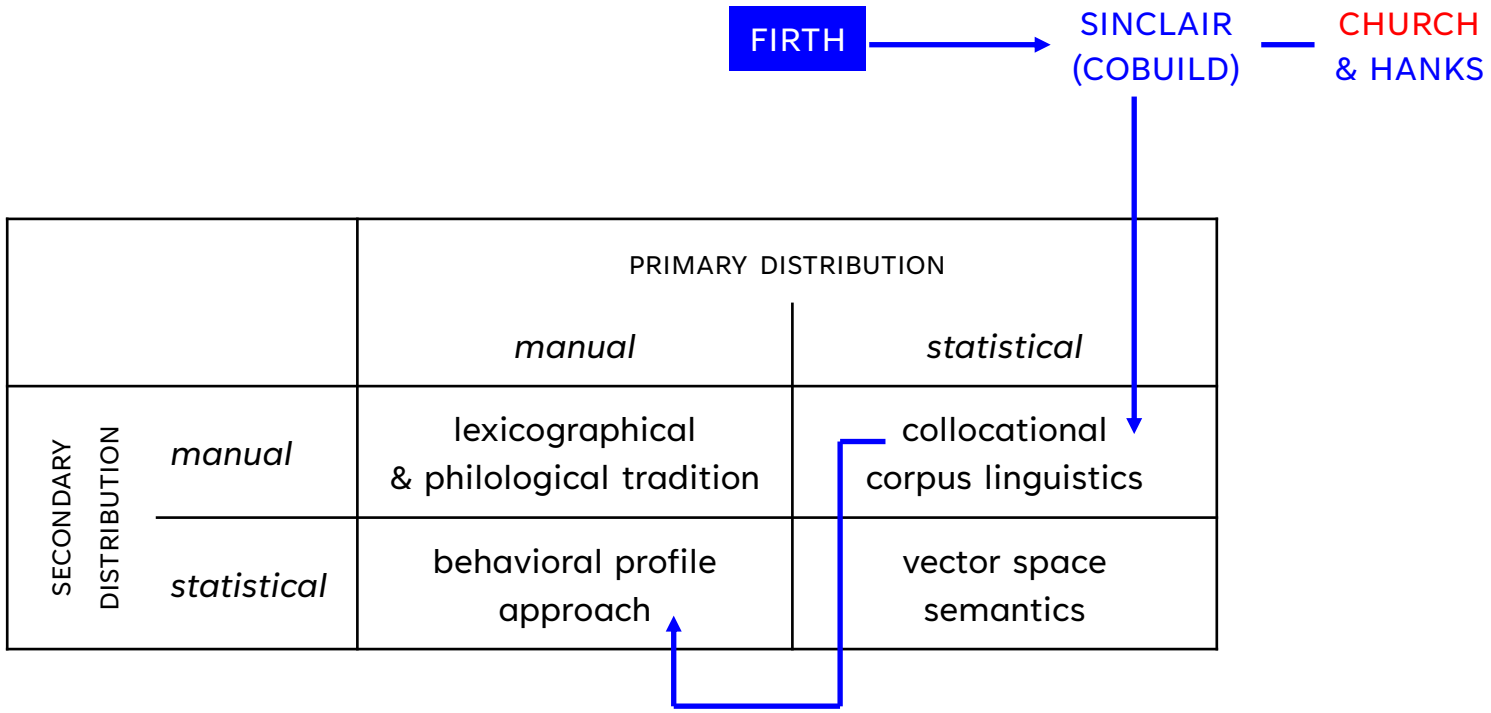
Sinclair 1987
Collins Cobuild Dictionary of Contemporary English

> Church & Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16: 22-29
>
> introduction of pmi as statistical method for identifying collocations
>
> ensuing development of tools like Sketch Engine

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH



3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH



3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

a computational trajectory:

1990s
statistical turn in computational linguistics

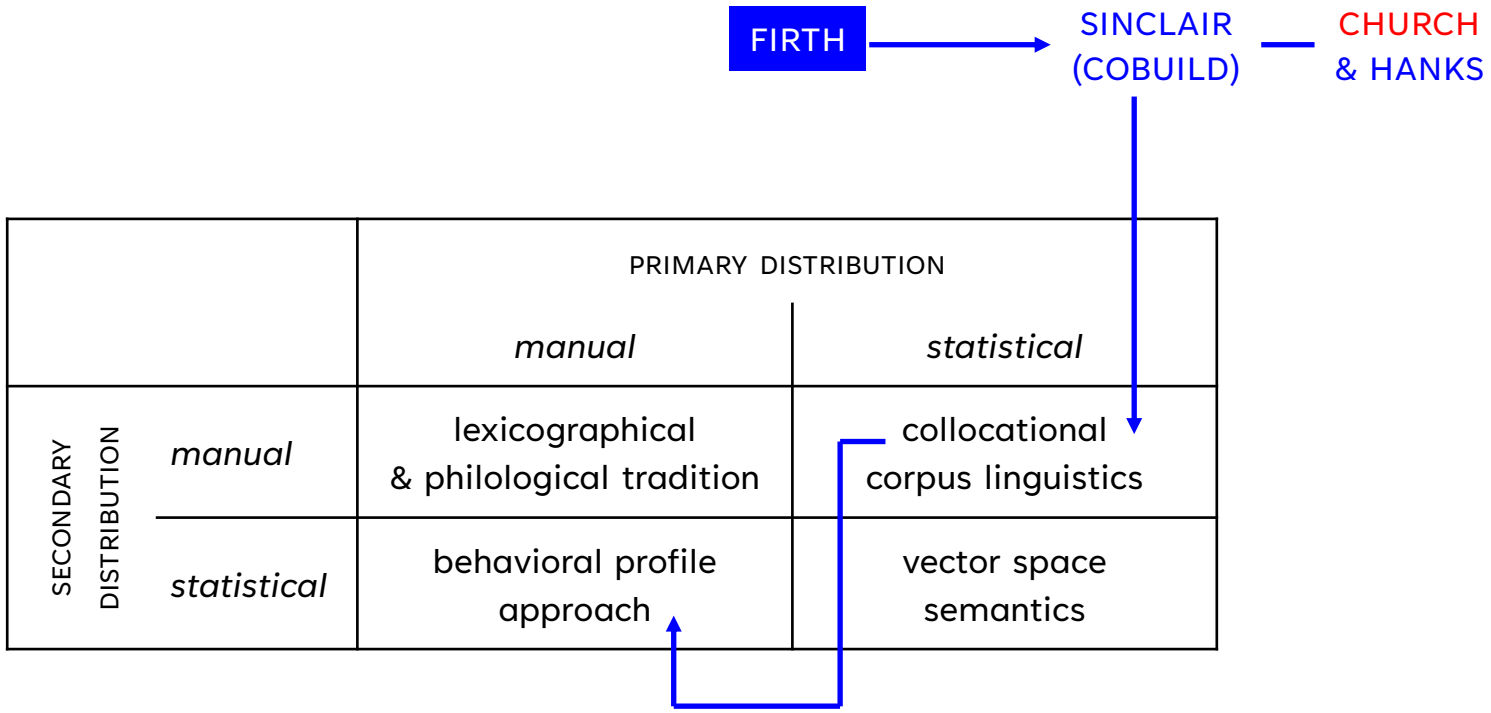
1993
ACL *Special Issue on Computational Linguistics Using Large Corpora*
Introduction by Church & Mercer mentions Firth, Harris
(and ‘1950s empiricism’) as forerunners
(but none of the papers does)

3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

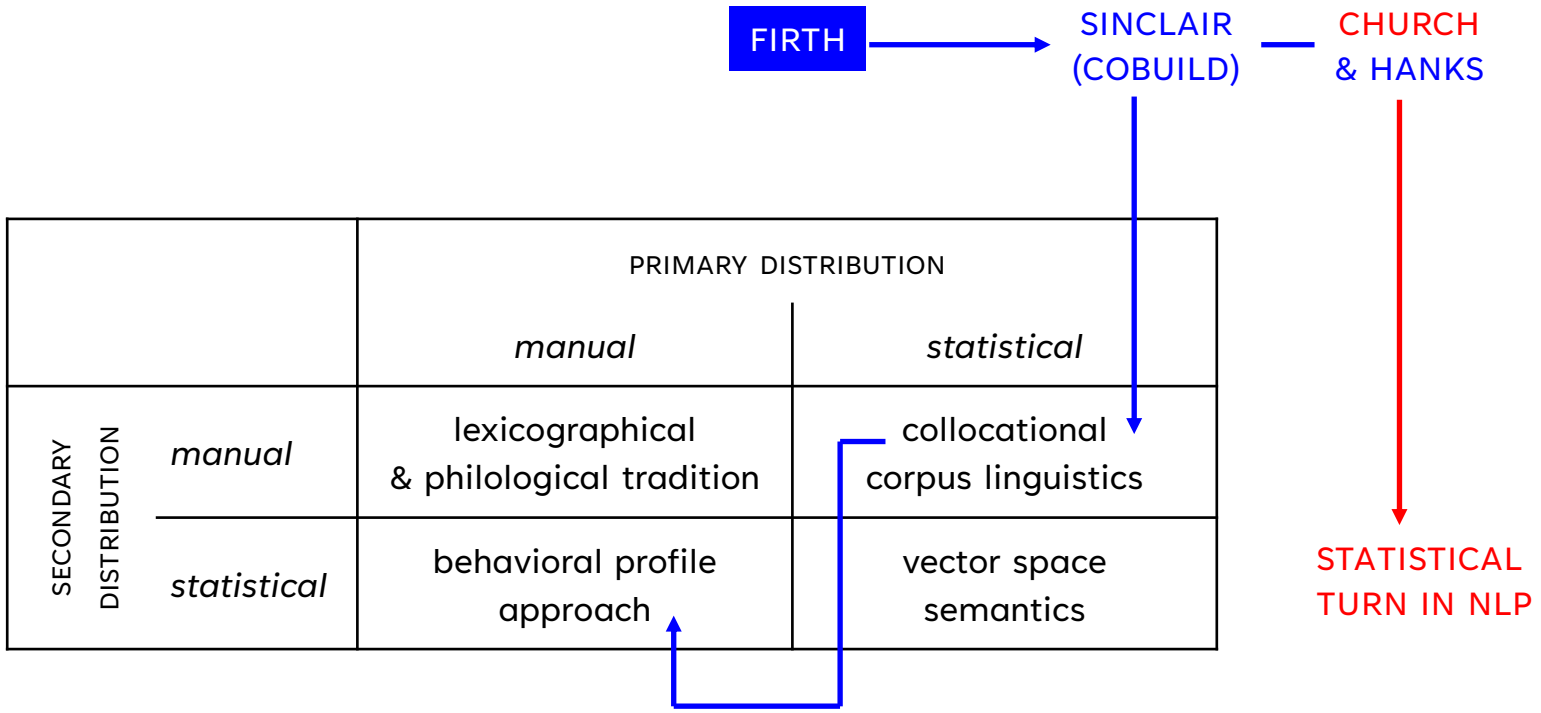
Table 6
Summary of two approaches to NLP.

	Rationalism	Empiricism
Well-known Advocates:	Chomsky, Minsky	Shannon, Skinner, Firth, Harris
Model:	Competence Model	Noisy Channel Model
Contexts of Interest:	Phrase Structure	N-grams
Goals:	All and Only	Minimize Prediction Error (Entropy)
	Explanatory	Descriptive
	Theoretical	Applied
Linguistic Generalizations:	Agreement and Wh-movement	Collocations and Word Associations
Parsing Strategies:	Principle-Based	Preference-Based
	CKY (Chart), ATNs, Unification	Forward–Backward, Inside–Outside
Applications:	Understanding	Recognition
	Who did what to whom	Noisy Channel Applications

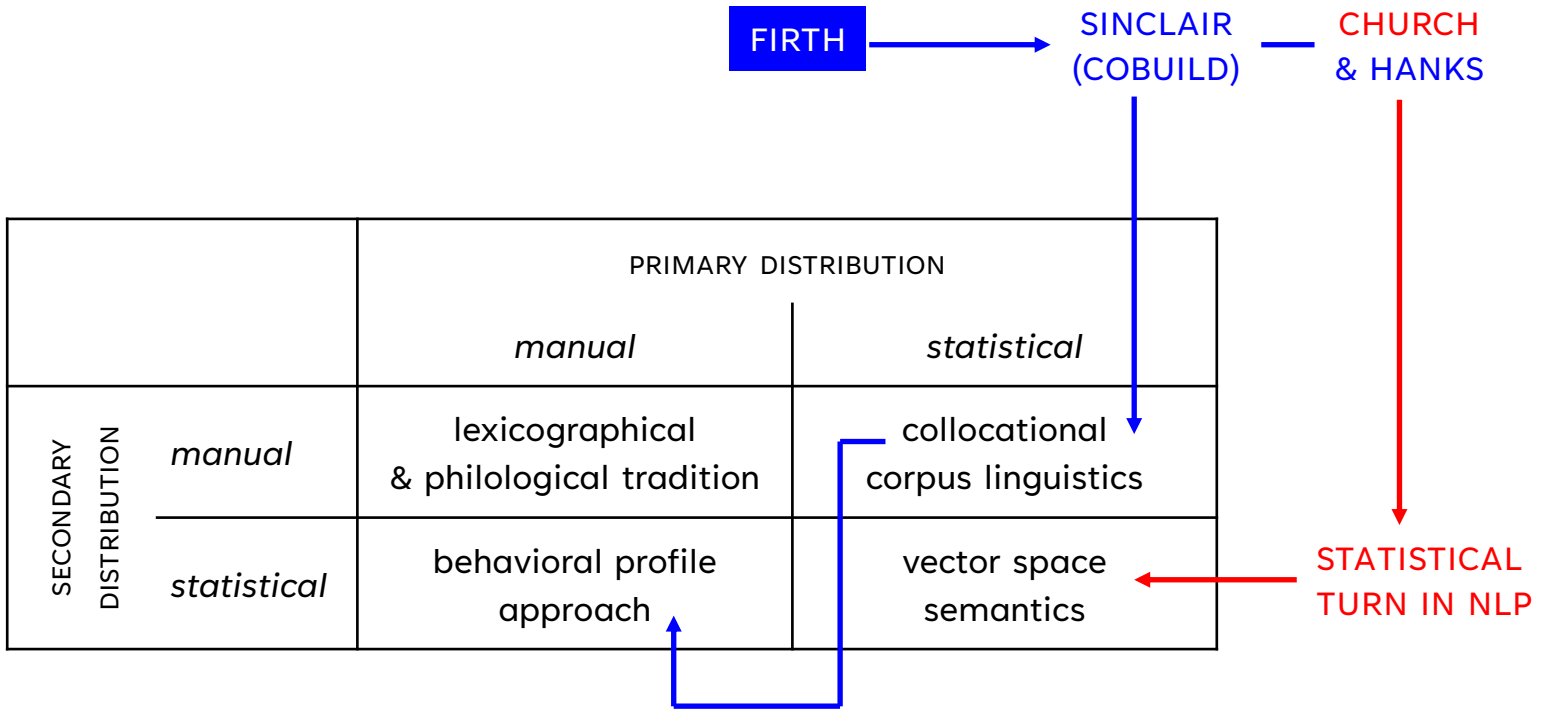
3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH



3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH



3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH



3 IMPACT ON DISTRIBUTIONAL SEMANTICS? - FIRTH

aftermath:
the linguistic and the computational Firthean lines meet when vector space modelling is incorporated in quantitative corpus linguistics

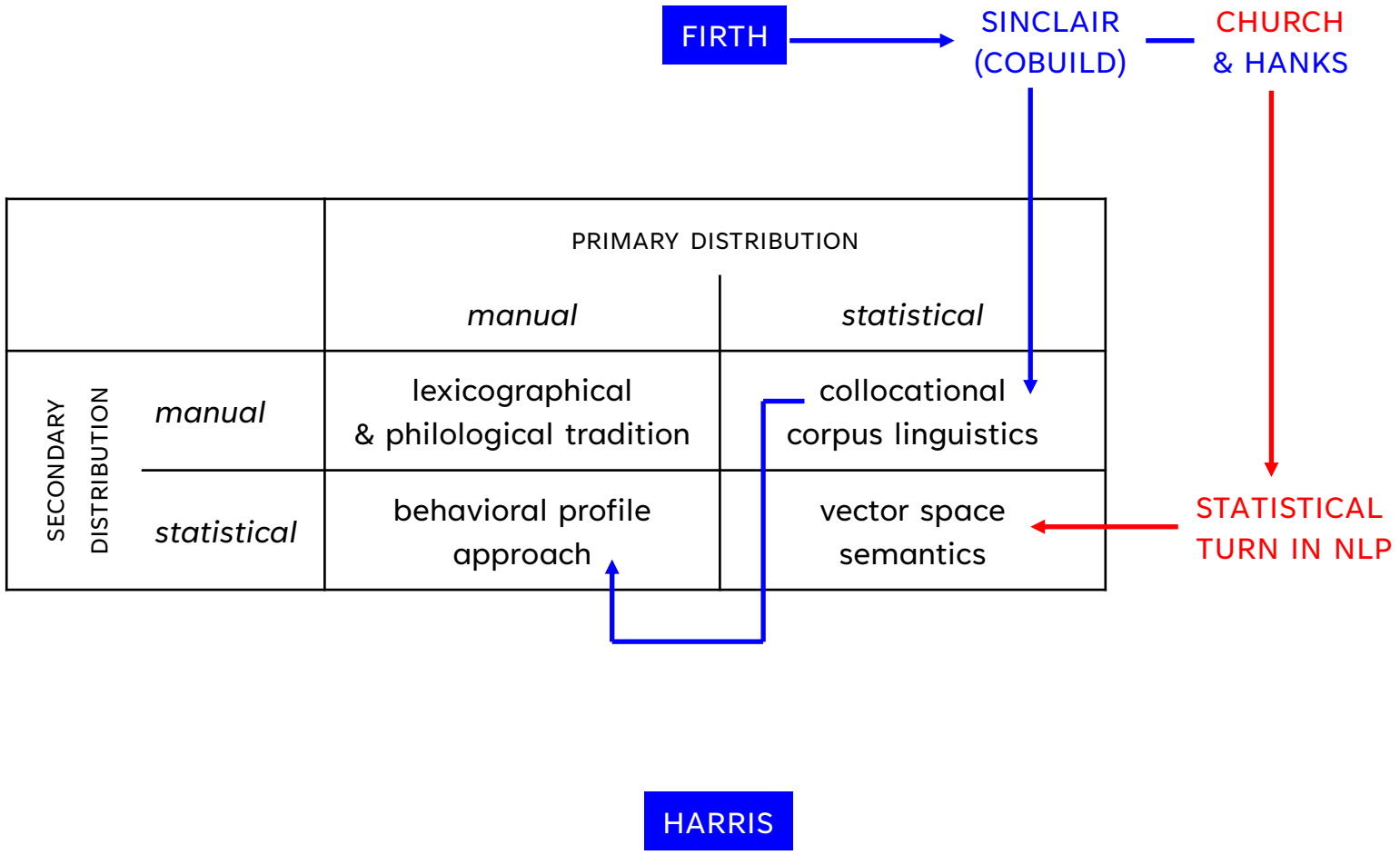
4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS

4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS

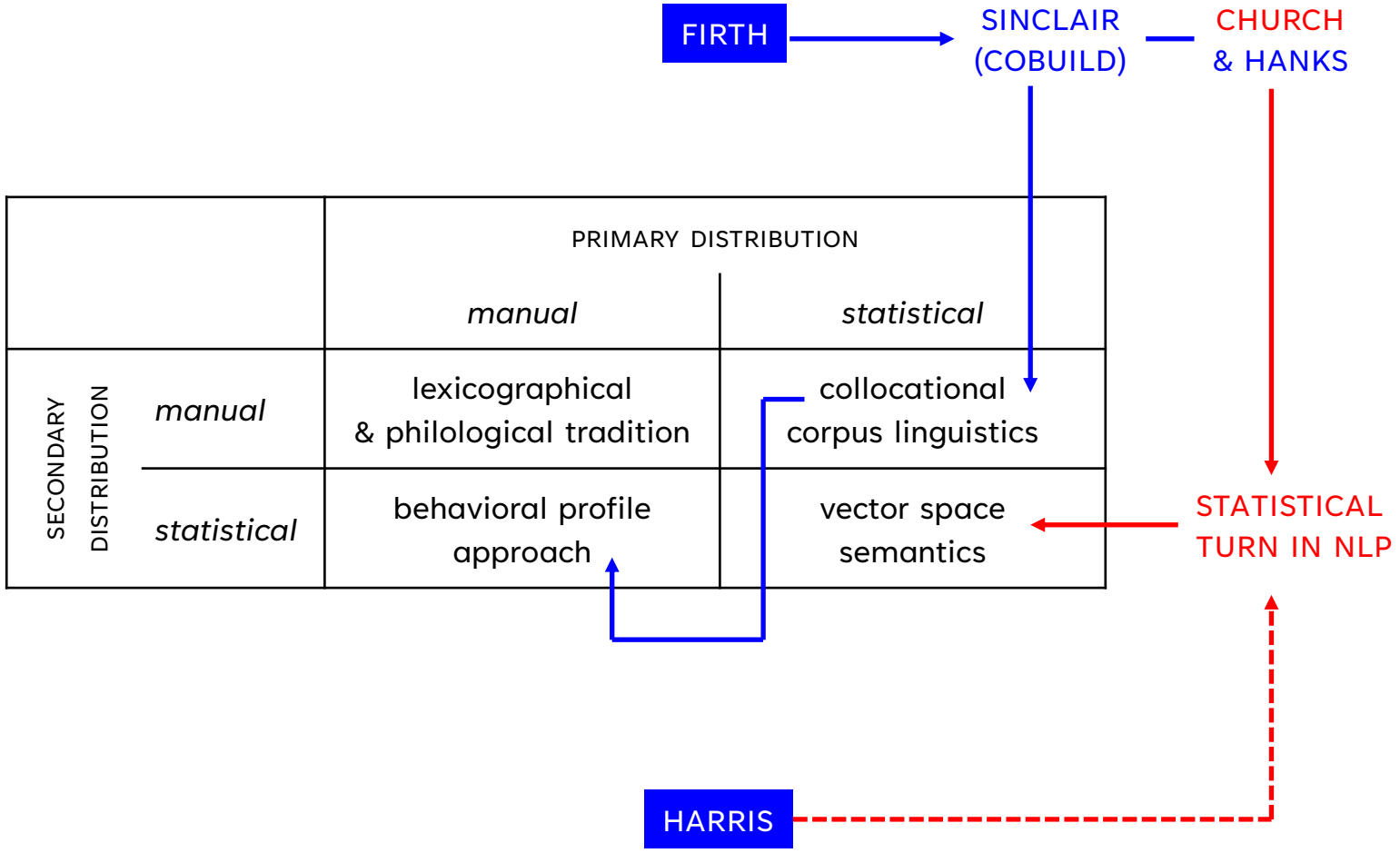
a computational trajectory:

in the switch from ‘rationalist’ to ‘empiricist’ computation (cp. Church & Mercer 1993), Harris is recognized as a forerunner of distributional and probabilistic modelling

4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS

a linguistic trajectory:

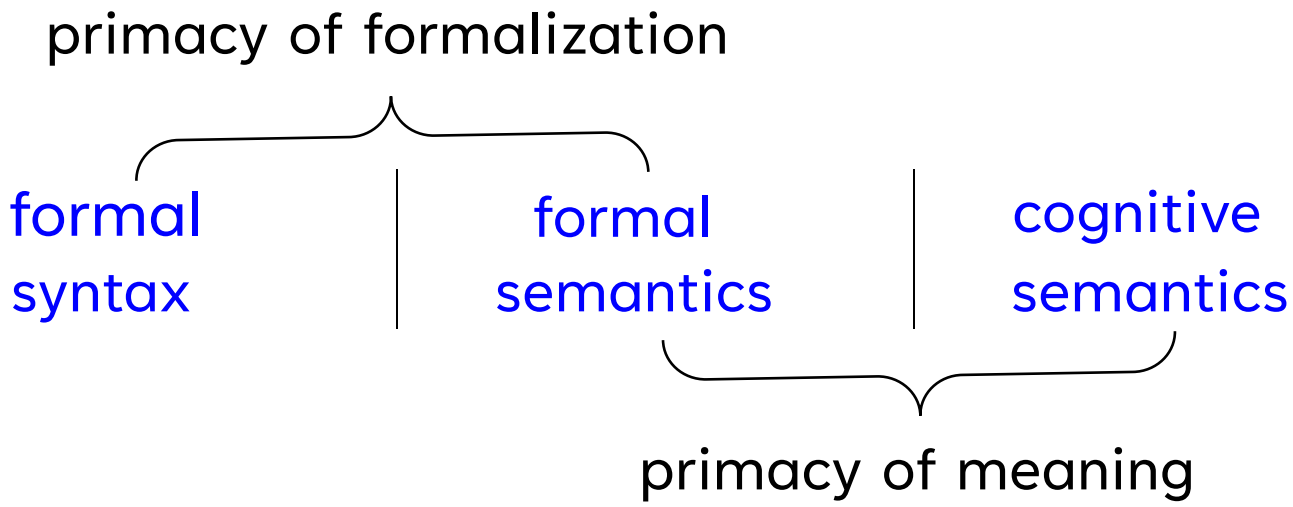
through his influence on Chomsky, Harris' minimization of semantics contributed indirectly to the emergence of cognitive and usage-based linguistics

4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS

first step:
Chomsky's changing attitude w.r.t. the semantic minimalism of his mentor Harris
1957 vs. 1965
→ the Linguistic Wars

4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS

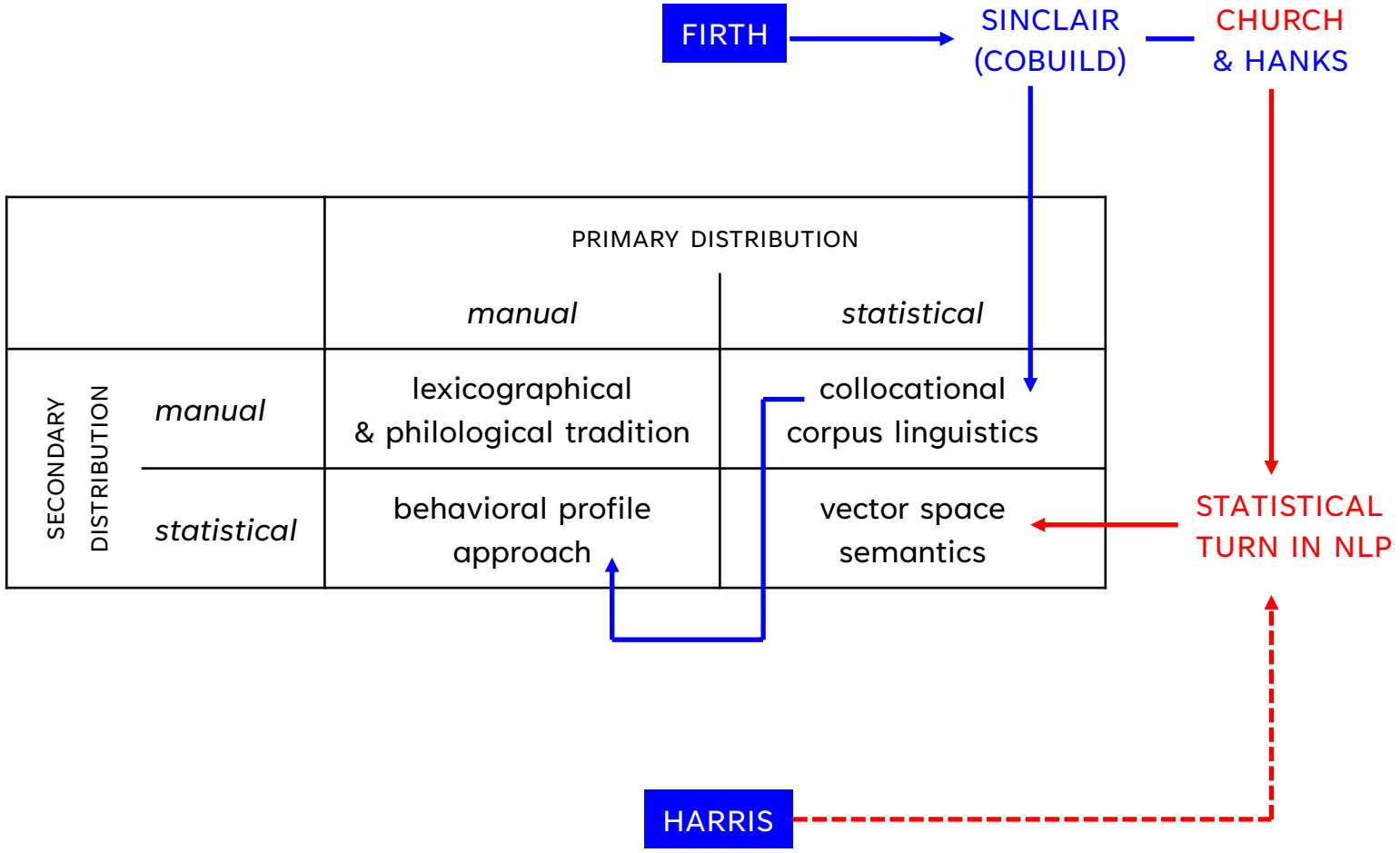
second step: consequences of the Linguistic Wars
a three-way split between
- formal grammarians
- formal semanticians
- cognitive semanticians



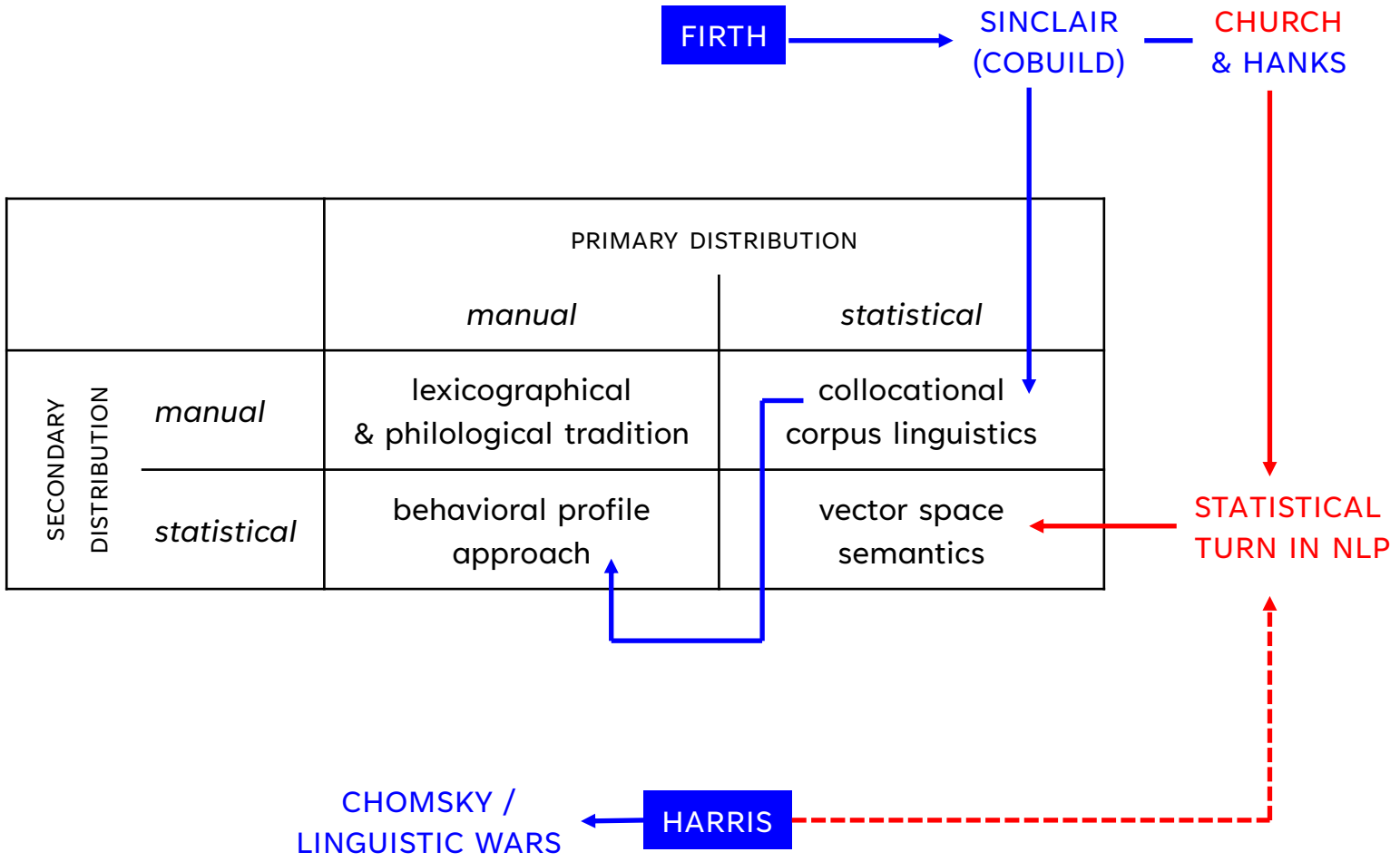
4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS

third step:
the ‘quantitative turn in cognitive linguistics’,
as before

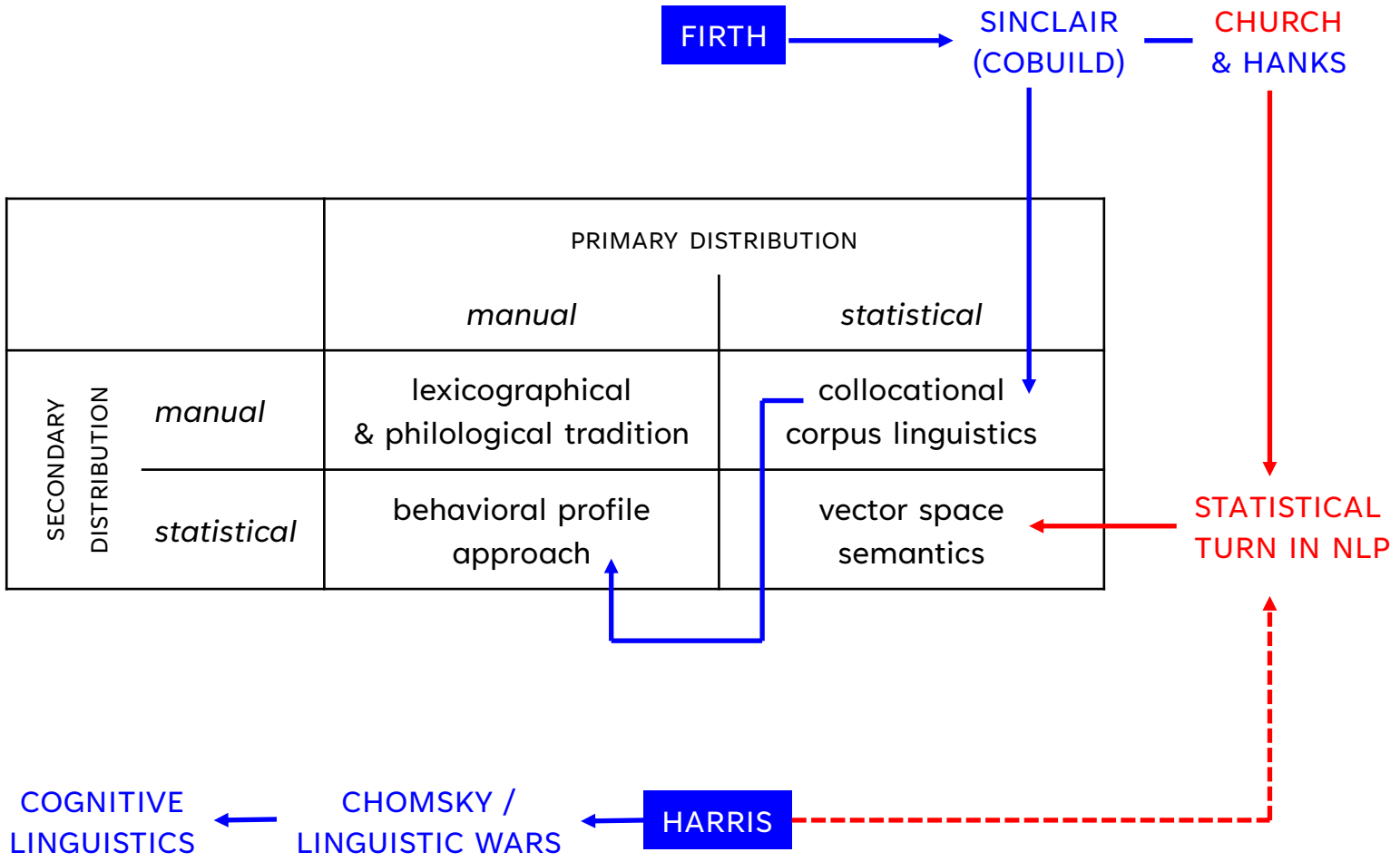
4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



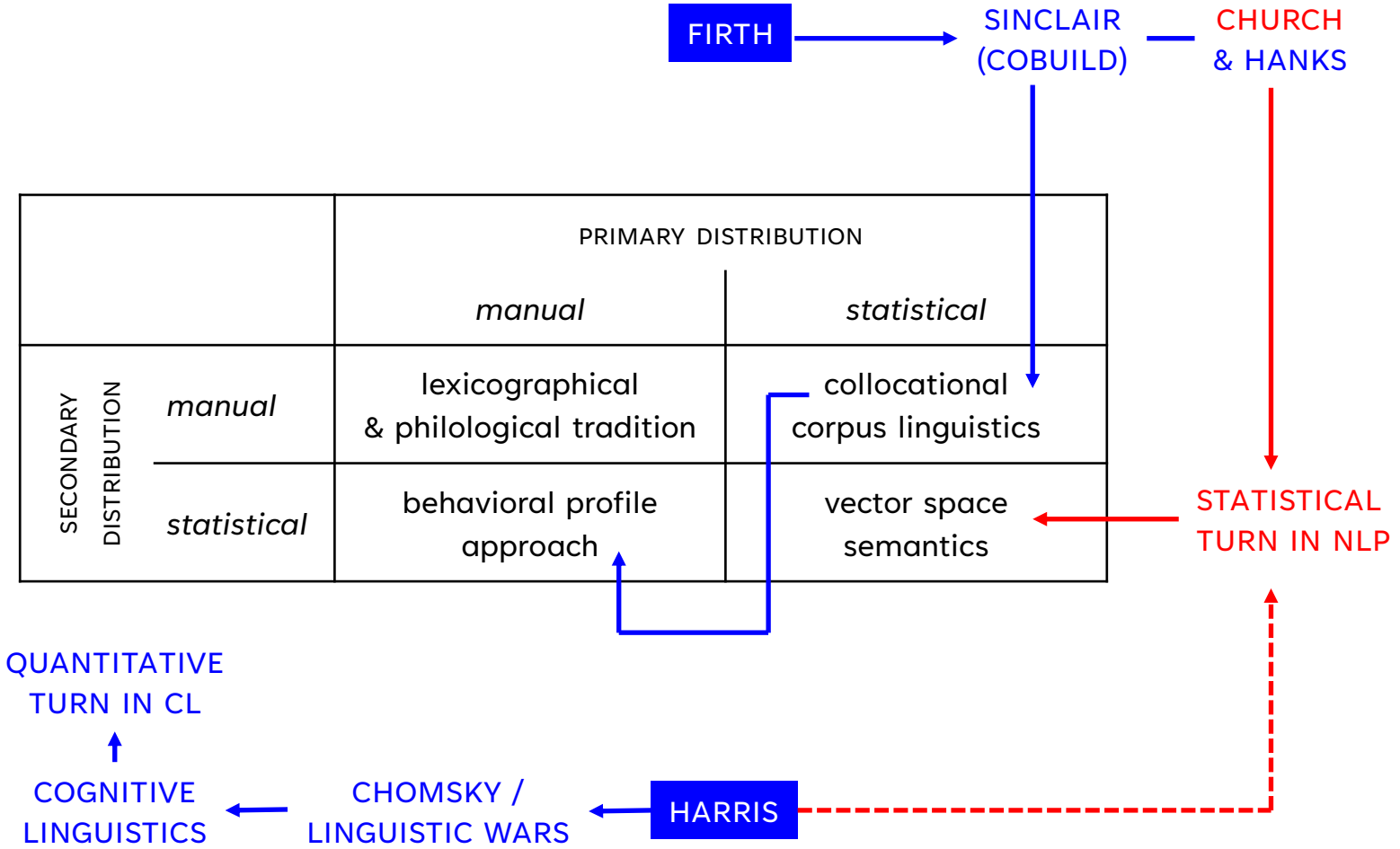
4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



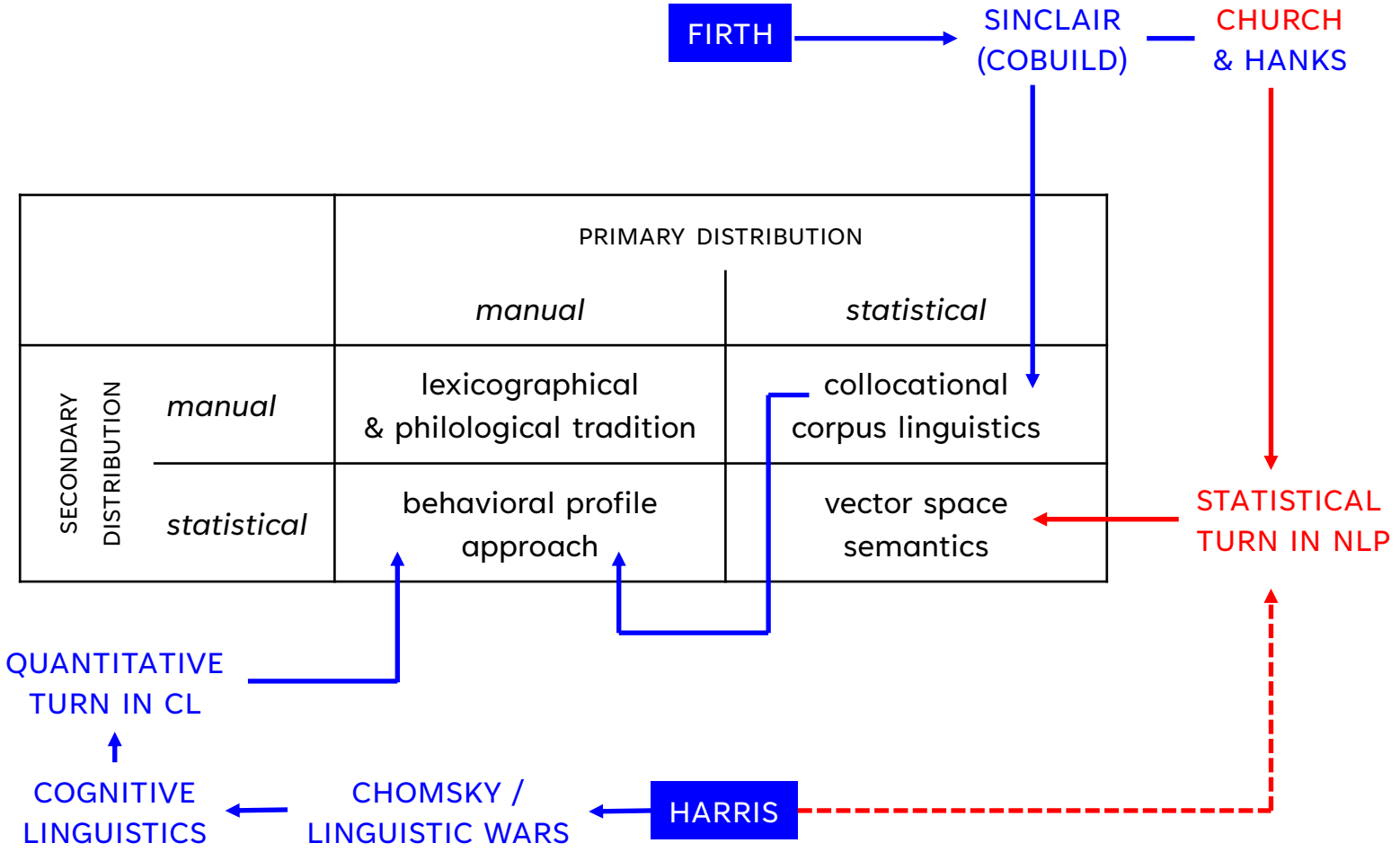
4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



4 IMPACT ON DISTRIBUTIONAL SEMANTICS? - HARRIS



CONCLUSIONS

CONCLUSIONS

in view of their actual ideas, references to H and F as founding figures of distributional semantics may be classified as largely ritual

any assessment of their actual influence needs to consider distributional semantics in its broadest sense, not just vector space semantics

the position of H and F differs in fundamental respects
→ what would that mean for contemporary semantics?

CONCLUSIONS

Harris
how can distribution tell us which structural elements exist in the language?
→ a heuristic approach

Firth
what does distribution tell us about the (social) meaning of expressions?
→ a hermeneutic approach

CONCLUSIONS

if, heuristically, distributional semantics involves identifying meanings, you need a thorough reflection on the notion of ‘meaning’

if, hermeneutically, distributional semantics involves providing information for interpretation, you need more transparency in the models

→ in both perspectives, there is work to be done

dirk.geeraerts@kuleuven

https://www.arts.kuleuven.be/ling/qlvl/people/pages/00003828

https://fdslive.oup.com/www.oup.com/academic/pdf/openaccess/9780198890676.pdf