

Historical linguistics data citation & reusability

...or data management principles for (historical) linguists

Blaine Billings
19 August 2025
ICHL27, Santiago de Chile



Focus of this presentation

Data design is an *integral* part of historical linguistic work

Focus of this presentation

Data design is an *integral* part of historical linguistic work

Data structure should reflect *design principles*

Focus of this presentation

Data design is an *integral* part of historical linguistic work

Data structure should reflect *design principles*

Data management is *not prohibitive* to the non-tech-savvy*

Focus of this presentation

Data design is an *integral* part of historical linguistic work

Data structure should reflect *design principles*

Data management is *not prohibitive* to the non-tech-savvy*

(*maybe just a little bit)

Outline

1. Data design principles
2. Historical linguistics data structure
3. Structuring historical linguistics data
4. Usability of database tools
5. Concluding thoughts

(← yes, they really are helpful!)

Part 1: Data design principles

Data design principles

1. **High cohesion** (Ku et al. 2006)
2. **Low coupling** (Ku et al. 2006)
3. **Atomicity** (Codd 1990)
4. **Reflectivity** (Gwinn 1992)

Data design principles

1. High cohesion

— *every table should have one (and only one) job* —

- **Example:**
 - Different tables for *each source*: in SCD, one table for Soravia (2007) Alas wordlist, another for Akbar et al. (1985) dictionary, another for Kreemer (1922) ethnography, ...
- **Issues this resolves:**
 - Source-restricted edits, such as normalizing different orthographies for each source, fixing off-counting page numbers, ...

Kreemer 1922

<ë>/<ö>/<ë>

<é>/<è>

<o>/<ò>

<oe>

Akbar et al. 1985

<e>

<ē>

<o>

<u>

Soravia 2007

<e>

<é>/<è>

<o>

<u>

Phonological

/ə/

/e/

/o/

/u/

Data design principles

1. High cohesion

— every table should have one (and only one) job —

2. **Low coupling**

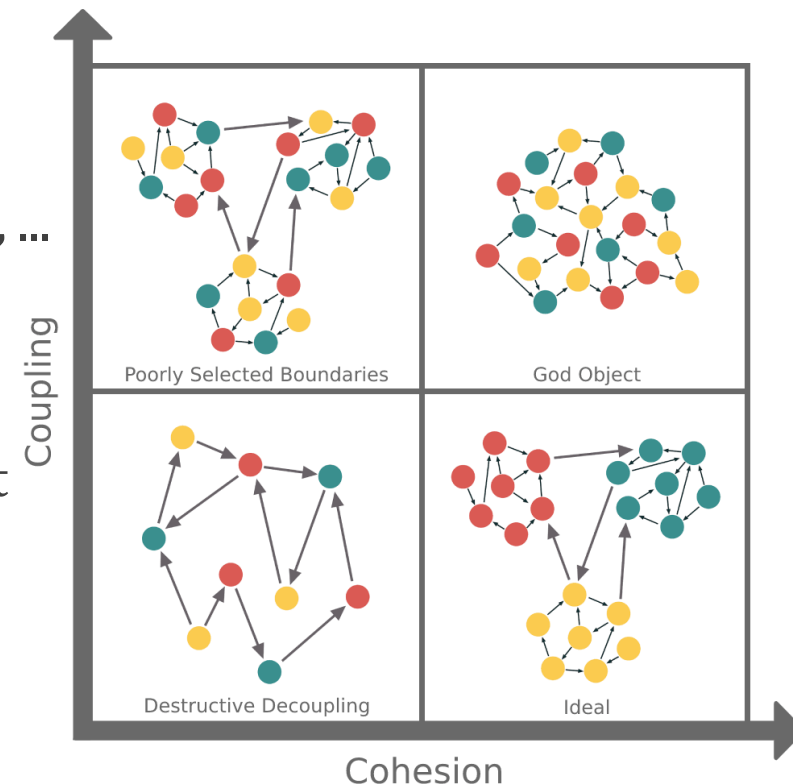
— **tables should only be connected where necessary** —

- **Example:**

- Source tables do not need to be connected to one another, nor do reconstructions, but sources with languages, descendant languages with ancestor languages, ...
- *Exception:* Loan relationships should be reflected in inter-language table connections

- **Issues this resolves:**

- Making one change means updating many tables, need not know everything about a data point before entering it, ...



Data design principles

1. High cohesion — *every table should have one (and only one) job* —
2. Low coupling — *tables should only be connected where necessary* —
3. **Atomicity** — ***each cell should only contain one part/type of data*** —
 - **Example:**
 - Orthography and phonological transcription should not appear in the same cell, a cell describing loan influence should not contain everything about that loan but rather be a row in another table, ...
 - **Issues this resolves:**
 - Bulk editing specific data fields, (regular expression) searches for content, filtering and viewing subsets of data, ...

(Loans) Sanskrit लशुन (laśuna) ‘garlic (allium sativum)’

⇒ Sanskrit | लशुन | laśuna | garlic

(Search) Sanskrit.*a\)

⇒ Sanskrit a\$

(Edit) Sanskrit शिक्षा (shikṣa) ‘punishment’

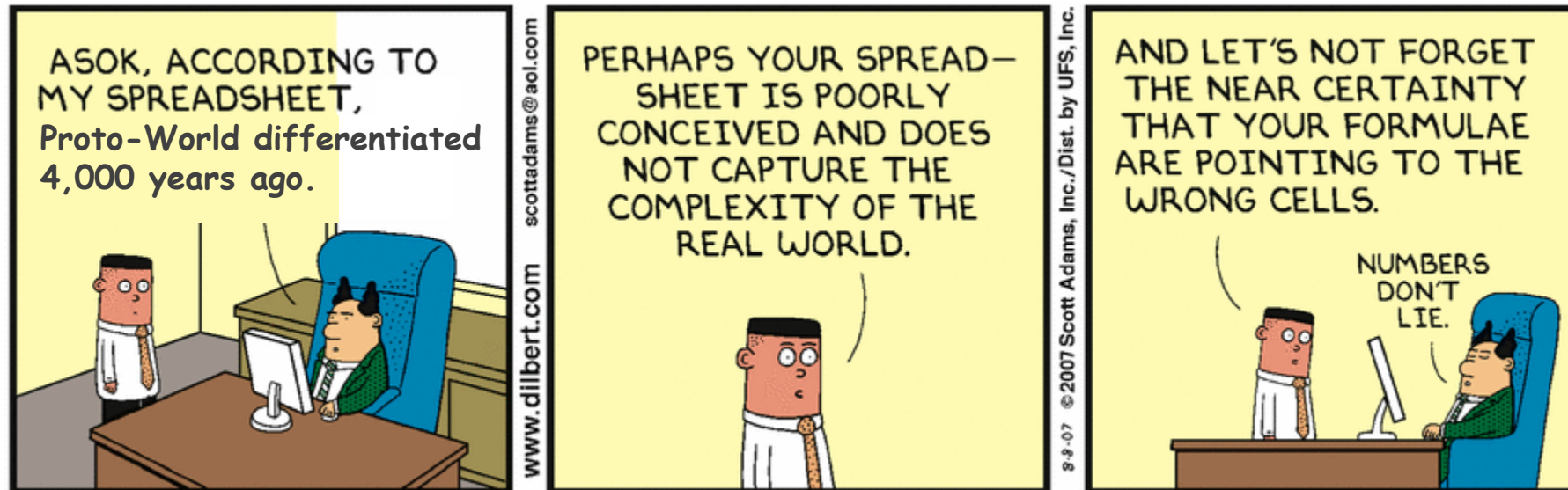
⇒ sh → ś without editing gloss

Data design principles

1. High cohesion — *every table should have one (and only one) job* —
2. Low coupling — *tables should only be connected where necessary* —
3. Atomicity — *each cell should only contain one part/type of data* —
4. **Reflectivity** — ***data structure should reflect data reality*** —

... brief intermission ...

Part 2: Historical linguistics data structure



Historical linguistics timeline

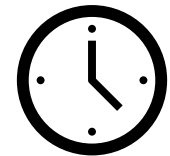
Basic historical linguistics timeline:

1. Language A has a “lexicon”
2. Lexical items in language A’s lexicon change over time
 - Changes can be in form, meaning, both
 - Changes can be innovations, losses,
 - Over time ancestor language A differentiates into descendant languages A-a, A-b, A-c, ...
3. Languages A-a, A-b, A-c, ... have their own “lexica”
4. ↻ Cycle repeats, languages develop

(*synchronic*)



(*diachronic*)



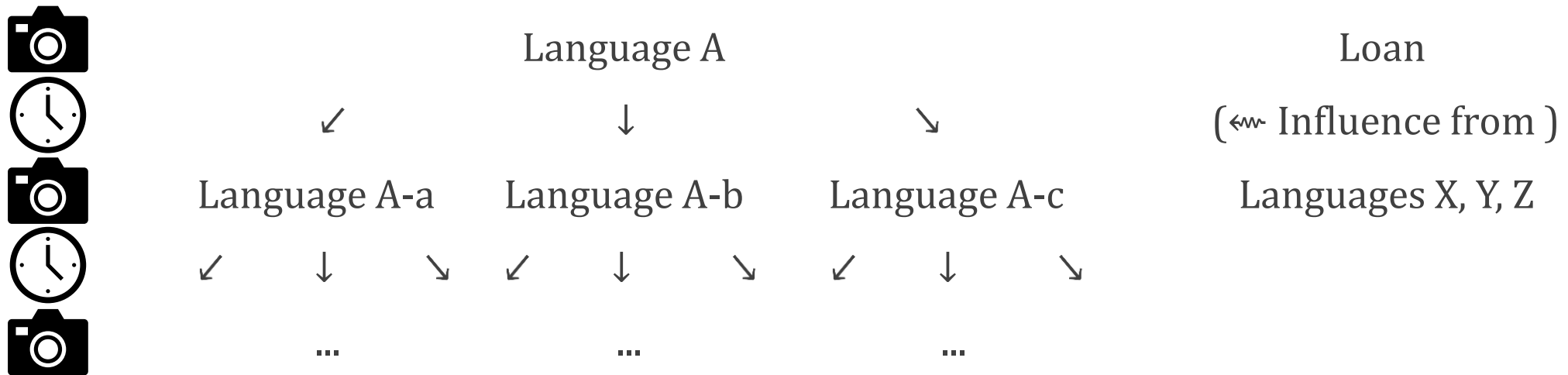
(*synchronic*)



Additionally: Different languages (and language *stages*) interact and lead to loan developments

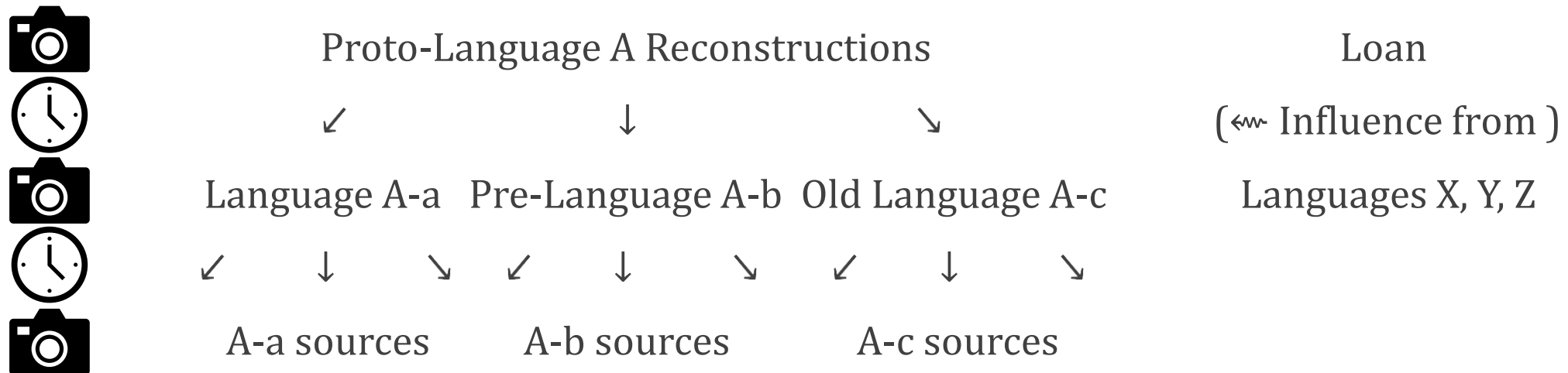
Historical linguistics timeline

Basic historical linguistics timeline can be broken down graphically:



Historical linguistics data

How does this look realistically, in terms of our data?



Sources:

- Dictionaries
- Recordings
- Passing remarks in other documents
- “Personal correspondence” that never really occurred but that we use to support our data...

Historical linguistics data

Often, this data is converted into a spreadsheet with one of the following formats...

Long form

Set #1	Lang A form	Lang A-a form	Lang A-b form	Lang A-c form	...	Loan form
Set #2	Lang A form	Lang A-a form	Lang A-b form	Lang A-c form	...	Loan form
...						

Tall form

Lang A form	Lang A-a form
Lang A form	Lang A-b form
Lang A form	Lang A-c form
...	

employees.txt - E:\work\temp\pipe

	1	2	3	4	5	6	7	8	9	10
1	DEPARTMENT_ID	LAST_NAME	FIRST_NAME	JOB_ID	SALARY	EMAIL	MANAGER_ID	COMMISSION_PCT	PH	
2	90	King	Steven	AD_PRES	24000	SKING			515.123.4567	100 1987-6-17
3	90	Kochhar	Neena	AD_VP	17000	NKOCHHAR	100		515.123.4568	101 1989-9-21
4	90	De Haan	Lex	AD_VP	17000	LDEHAAN	100		515.123.4569	102 1993-1-13
5	60	Hunold	Alexander	IT_PROG	9000	AHUNOLD	102		590.423.4567	103 1990-1-3
6	60	Ernst	Bruce	IT_PROG	6000	BERNST	103		590.423.4568	104 1991-5-21
7	60	Austin	David	IT_PROG	4800	DAUSTIN	103		590.423.4569	105 1997-6-25
8	60	Pataballa	Valli	IT_PROG	4800	VPATABAL	103		590.423.4560	106 1998-2-5
9	60	Lorentz	Diana	IT_PROG	4200	DLORENTZ	103		590.423.5567	107 1999-2-7
10	100	Greenberg	Nancy	FI_MGR	12000	NGREENBE	101		515.124.4569	108 1994-8-17
11	100	Faviet	Daniel	FI_ACCOUNT	9000	DFAVIET	108		515.124.4169	109 1994-8-16
12	100	Chen	John	FI_ACCOUNT	8200	JCHEN	108		515.124.4269	110 1997-9-28
13	100	Sciarra	Ismael	FI_ACCOUNT	7700	ISCIARRA	108		515.124.4369	111 1997-9-30
14	100	Urman	Jose Manuel	FI_ACCOUNT	7800	JMURMAN	108		515.124.4469	112 1998-3-7
15	100	Popp	Luis	FI_ACCOUNT	6900	LPOPP	108		515.124.4567	113 1999-12-7
16	30	Raphaely	Den	PU_MAN	11000	DRAPHEAL	100		515.127.4561	114 1994-12-7
17	30	Khoo	Alexander	PU_CLERK	3100	AKHOO	114		515.127.4562	115 1995-5-18
18	30	Baida	Shelli	PU_CLERK	2900	SBAIDA	114		515.127.4563	116 1997-12-24
19	30	Tobias	Sigal	PU_CLERK	2800	STOBIAS	114		515.127.4564	117 1997-7-24
20	30	Himuro	Guy	PU_CLERK	2600	GHIMURO	114		515.127.4565	118 1998-11-15
21	30	Colmenares	Karen	PU_CLERK	2500	KCOLMENA	114		515.127.4566	119 1999-8-10
22	50	Weiss	Matthew	ST_MAN	8000	MWEISS	100		650.123.1234	120 1996-7-18
23	50	Fripp	Adam	ST_MAN	8200	AFRIPP	100		650.123.2234	121 1997-4-10
24	50	Kaufling	Payam	ST_MAN	7900	PKAUFLIN	100		650.123.3234	122 1995-5-1
25	50	Vollman	Shanta	ST MAN	6500	SVOLLMAN	100		650.123.4234	123 1997-10-10

employees.txt - E:\work\temp\pipe

		1	2	3	4	5
1	DEPARTMENT_ID	LAST_NAME	FIRST_NAME	JOB_ID	SAL	
2	'90'	'King'	'Steven'	'AD_PRES'	'24000'	'SKING'
3	'90'	'Kochhar'	'Neena'	'AD_VP'	'17000'	'NKOCHHAR'
4	'90'	'De Haan'	'Lex'	'AD_VP'	'17000'	'LDEHAAN'
5	'60'	'Hunold'	'Alexander'	'IT_PROG'	'9000'	'AHUNOLD'
6	'60'	'Ernst'	'Bruce'	'IT_PROG'	'6000'	'BERNST'
7	'60'	'Austin'	'David'	'IT_PROG'	'4800'	'DAUSTIN'
8	'60'	'Pataballa'	'Valli'	'IT_PROG'	'4800'	'VPATABAL'
9	'60'	'Lorentz'	'Diana'	'IT_PROG'	'4200'	'DLORENTZ'
10	'100'	'Greenberg'	'Nancy'	'FI_MGR'	'12000'	'NGREENBE'
11	'100'	'Faviet'	'Daniel'	'FI_ACCOUNT'	'9000'	'DFAVIET'
12	'100'	'Chen'	'John'	'FI_ACCOUNT'	'8200'	'JCHEN'
13	'100'	'Sciarra'	'Ismael'	'FI_ACCOUNT'	'7700'	'ISCIARRA'
14	'100'	'Urman'	'Jose Manuel'	'FI_ACCOUNT'	'7800'	'JMURM'
15	'100'	'Popp'	'Luis'	'FI_ACCOUNT'	'6900'	'LPOPP'
16	'30'	'Raphaely'	'Den'	'PU_MAN'	'11000'	'DRAPHEAL'
17	'30'	'Khoo'	'Alexander'	'PU_CLERK'	'3100'	'AKHOO'
18	'30'	'Baida'	'Shelli'	'PU_CLERK'	'2900'	'SBIDA'
19	'30'	'Tobias'	'Sigal'	'PU_CLERK'	'2800'	'STOBIAS'
20	'30'	'Himuro'	'Guy'	'PU_CLERK'	'2600'	'GHIMURO'
21	'30'	'Colmenares'	'Karen'	'PU_CLERK'	'2500'	'KCOLMENA'
22	'50'	'Weiss'	'Matthew'	'ST_MAN'	'8000'	'MWEISS'
23	'50'	'Fripp'	'Adam'	'ST_MAN'	'8200'	'AFRIPP'
24	'50'	'Kaufling'	'Payam'	'ST_MAN'	'7900'	'PKAUFLIN'
25	'50'	'Vollman'	'Shanta'	'ST_MAN'	'6500'	'SVOLLMAN'

KEY	Double GW		Possible Double GW		Double GW and Blank GW				Unrescheduled Blank	Possible Future Postponements					
	Blank GW		Possible Blank GW		Double GW and Possible Blank GW										
Gameweek:	32	33		34	35	36		37	38	Gameweek:	33	Gameweek:	37		
ARS	vs	BHA	sou	che	MUN	whu	LEE	tot	new	EVE	ARS		ARS		
AVL	vs	TOT			lei	NOR	CRY?	bur	LIV	CRY?	BUR	mci	AVL	vs	CRY
BRE	vs	WHU	wat		TOT	mun	SOU		eve	LEE	BRE		BRE		
BHA	vs	ars	tot	mci	SOU	wol	MUN		lee	WHU	BHA		BHA		
BUR	vs	nor	whu	SOU	WOL	wat	AVL		tot	avl	NEW	BUR		BUR	
CHE	vs	sou	ARS		WHU	mun?	eve	WOL	lee	mun?	LEI	WAT	CHE	vs	mun
CRY	vs	lei	new		LEE	sou	avl?	WAT		avl?	eve	MUN	CRY	vs	avl
EVE	vs	MUN	LEI		liv	CHE	lei	wat	BRE	CRY	ars	EVE		EVE	
LEE	vs	wat			cry	MCI	ars	CHE	BHA		bre	LEE		LEE	
LEI	vs	CRY	new	eve	AVL	tot	EVE	NOR	wat	che	SOU	LEI		LEI	
LIV	vs	mci	MUN		EVE	new	TOT	avl	sou		WOL	LIV	vs	sou	
MCI	vs	LIV	BHA		WAT	lee	NEW	whu/wol?	whu?	wol?	AVL	MCI	vs	whu	
MUN	vs	eve	NOR	liv	ars	CHE?	BRE	bha	CHE?		cry	MUN	vs	CHE	
NEW	vs	WOL	LEI	CRY	nor	LIV	mci		ARS		bur	NEW		NEW	
NOR	vs	BUR	mun		NEW	avl	WHU	lei	wol		TOT	NOR		NOR	
SOU	vs	CHE	ARS	bur	bha	CRY	bre		LIV		lei	SOU	vs	LIV	
TOT	vs	avl	BHA		bre	LEI	liv	ARS	BUR		nor	TOT		TOT	
WAT	vs	LEE	BRE		mci	BUR	cry	EVE	LEI		che	WAT		WAT	
WHU	vs	bre	BUR		che	ARS	nor	MCI?	MCI?		bha	WHU	vs	MCI	
WOL	vs	new			bur	BHA	che	MCI?	NOR	MCI?	liv	WOL	vs		
Fixture formatting	1	2	3	4	5	6	7	8	9	10	11	12	13	Created by @BenCrellin in partnership with Fantasy Football Hub	
	Easiest - best vs worst					Average				Toughest - worst vs best					

employees.txt - E:\work\temp\pipe

		1	2	3	4	5
1	DEPARTMENT_ID	LAST_NAME	FIRST_NAME	JOB_ID	SAL	
2	'90'	'King'	'Steven'	'AD_PRES'	'24000'	'SKING'
3	'90'	'Kochhar'	'Neena'	'AD_VP'	'17000'	'NKOCHHAR'
4	'90'	'De Haan'	'Lex'	'AD_VP'	'17000'	'LDEHAAN'
5	'60'	'Hunold'	'Alexander'	'IT_PROG'	'9000'	'AHUNOLD'
6	'60'	'Ernst'	'Bruce'	'IT_PROG'	'6000'	'BERNST'
7	'60'	'Austin'	'David'	'IT_PROG'	'4800'	'DAUSTIN'
8	'60'	'Pataballa'	'Valli'	'IT_PROG'	'4800'	'VPATABAL'
9	'60'	'Lorentz'	'Diana'	'IT_PROG'	'4200'	'DLORENTZ'
10	'100'	'Greenberg'	'Nancy'	'FI_MGR'	'12000'	'NGREENBE'
11	'100'	'Faviet'	'Daniel'	'FI_ACCOUNT'	'9000'	'DFAVIET'
12	'100'	'Chen'	'John'	'FI_ACCOUNT'	'8200'	'JCHEN'

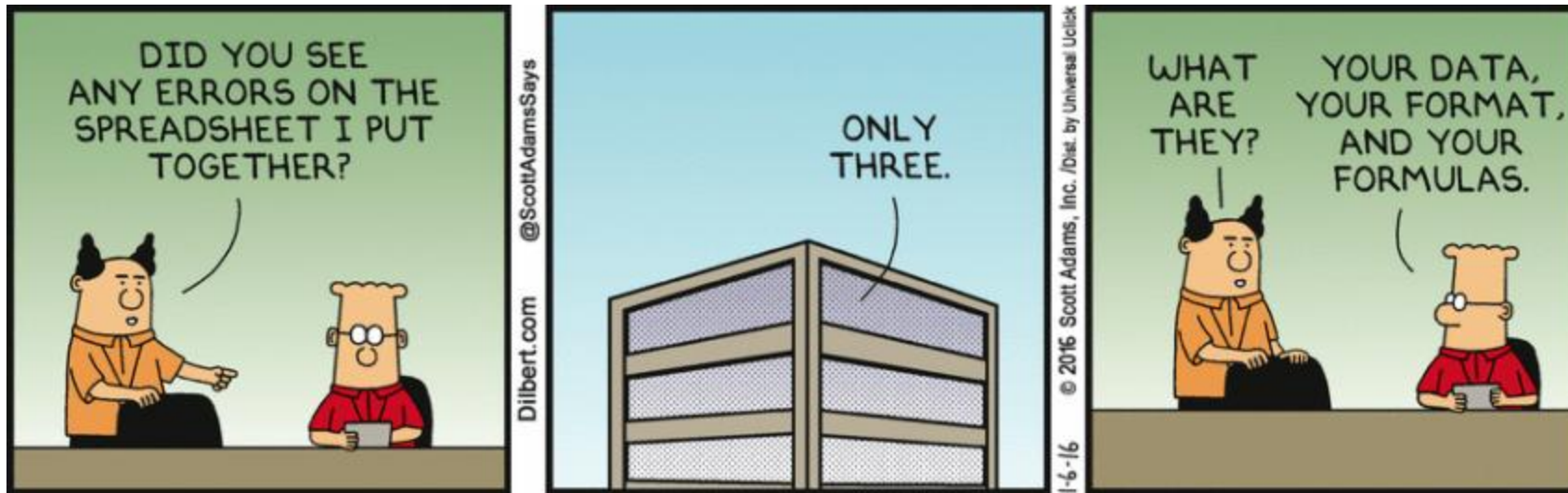
KEY	Double GW	Possible Double GW	Double GW and Blank GW		Unrescheduled Blank	Possible Future Postponements
	Blank GW	Possible Blank GW	Double GW and Possible Blank GW			
Gameweek:	32	33	34	35	36	37
ARS	vs BHA	sou che	MUN	whu	LEE tot	new EVE
AVL	vs TOT		lei	NOR CRY?	bur LIV	CRY? BUR mci
BRE	vs WHU	wat	TOT	mun	SOU	eve LEE
BHA	vs ars	tot mci	SOU	wol	MUN	lee WHU
BUR	vs nor	whu SOU	WOL	wat	AVL	tot avl NEW
CHE	vs sou	ARS	WHU mun?	eve	WOL lee	mun? LEI WAT
CRY	vs lei	new	LEE	sou avl?	WAT	avl? eve MUN
EVE	vs MUN	LEI	liv	CHE	lei wat	BRE CRY ars
LEE	vs wat		cry	MCI	ars CHE	BHA bre
LEI	vs CRY	new eve	AVL	tot	EVE NOR	wat che SOU
LIV	vs mci	MUN	EVE	new	TOT avl	sou WOL

Gameweek:	33		
ARS			
AVL			
BRE			
BHA			
BUR			
CHE	vs mun		
CRY	vs avl		
EVE			
LEE			
LEI			
LIV	vs sou		

Gameweek:	37		
ARS			
AVL	vs CRY		
BRE			
BHA			
BUR			
CHE	vs mun		
CRY	vs avl		
EVE			
LEE			
LEI			
LIV	vs sou		

A	R	S	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	P
Category	Hi	Meaning	PMP: 1656	Gayo: 357	Al: 532	Ka: 631	Da: 964	PNB: 1094	Simal: 587	To: 720	An: 640	Ma: 630	PSB: 776	PCSB: 825	PB: 851	Ha: 203	Le: 269	Sim: 792	PNBI: 726	Sig: 277	Nias: 743	P
Body Parts	2.1	Head, Upstream	*qulu	ulu		ulu	ulu	*ulu	ulu	ulu	ulu	ulu	*ulu	*ulu	*ulu	ulu	ulu	ulu	*ulu	ulu	ulu	*ulu
	2.2	Head			takal	takal	takal	*takal														
	3.1	Face	*away		awe	ayo	abe	*awe	awei	ae			*ae	*away	*away							
	3.2	Face														wangon	bangon	bobangon	*banjon		angango	*anja
	5	Skull														lalar	lalar	lalar	*lalar			
	6.1	Hair	*buhək	(w)uk	buk	buk	buk	*buk		obuk	obuk	obuk	*obuk	*obuk	*buk	wok	bo	bu'o?	*buk	bu	bu	*bu
	6.2	White Head Hair	*quban		uban	uban		*uban							*uban			ufan	*uban	zufa	hufa	*hut
	7	Bald	*palpal	pepal	papal	pal-pal		*pal-pal	pal-pal	pal-pal	pal-pal	pal-pal	*pal-pal	*pal-pal	*pal-pal							
	8.1	Crown	*bubun	ubun-ubun	bubun		babon	*bubu-an	sa(li)mbubu	(s)ambubu	ambubu	ambubu	*ambubu	*ambubu	*bubu	wowot	bufus	bubuh	*bubu-an			
	8.3	Fontanel	*ambut			emput-emput		*emput	ombut	ombut-ombut	ombut-ombut	ombut-ombut	*ombut	*ombut	*ambut							
	8.4	Fontanel	*ubun		ubun			*ubun							*ubun							
	9.1	Ear	*kupin	cuping	cuping	cuping	coping	*cupin		suping	suping	suping	*supin	*supin	*cupin							
	9.2	Ear			pinggel	penggel		*pinggal	pinggol	pinggol	pinggol	pinggol	*pinggol	*pinggol	*pinggal		eggel	eggel				
	9.3	Ear														oyok	xuyuk	uyuh	*kuyuk	kukuyu		*ku-
	9.4	Ear	*taliŋa																		taliŋa	*tali
	11.1	Eye	*mata	mata	mate	mata	mata	*mata	mata	mata	mata	mata	*mata	*mata	*mata	mata	mata	mata	*mata	mata	mata	*ma-
	11.2	Sty on the Eye	*mis.mis							halimismison	halimismis	halimismis	*hali-mis-mis	*hali-mis-mis	*kali-mis-mis							
	11.3	Mucus in the Ey	*mutaq																		muti	*mu
	11.4	Cataract of the E	*putik	putik																		
	11.5	Eye, To Look																		ufe	ufe	*ube
	11.6	To Blink	*kizap		kidep	kidep	singkerdep	*kidep	hidop					*hidop	*kidep							
	14	Eyebrow		seliben	seliben	seliben	seleben	*seliben	salibon	salibon	salibon	salibon	*salibon	*salibon	*saliben							
	16.1	Tear	*luhaq	luh	luh	luh	luh	*luh	luh	luh	luh	luh	*luh	*luh	*i-luh					elo		*elo
	16.2	Tear	*wahiR ni mata																		weh mata	*weh mata
	18	Nose	*iŋuŋ	iung	igung	igung	egung	*igun	igung	igung	igung	igung	*igun	*iguŋ	*iguŋ	ixong	ixong	ihung	*ixun	ixu	ixu	*ixu
	20.1	Mucus		ungi					sungui					*suŋuy	*suŋuy							
	20.2	Mucus			imen	imen	emen	*iman		mon-mon	mon-mon	mon-mon	*mon-mon	*mon-mon	*imən							
	20.3	Mucus	*guhuR													ingol		ingol	*inol		ingo	*ingo
	21	Cheek	*pihpih	pipi			pipi	*pipi	pipi	pipi			*pipi	*pipi	*pipi						fifi	*fifi
	21	Cheek	*hasan														asan	asan	*asan		boʔi	boʔi
	21	Cheek																				
	21	Cheek			kurum	kurum	kurum	*kurum	huyum	hurum	hurum	urum	*hurum	*hurum	*kurum							
	22	Mouth	*baqbaq	(b)abah	babah	babah	babah	*babah	babah	baba	baba	baba	*baba	*babah	*babah	wakba	bakba	baʔba	*bakbak	bafa	baʔa	*bab

Part 3: Structuring historical linguistic data



Historical linguistics data structuring

This structure introduces a **number of issues**:

- It ***does not*** reflect language reality
- It ***does not*** abide by design principles
- It ***often*** must be transformed for reuse
- It is ***often*** not immediately understandable

So, how ***might*** we structure data?

Historical linguistics data structuring

So, how *might* we structure data?

Source

Entry

Definition

Page Number

Historical linguistics data structuring

So, how *might* we structure data?

AlsKr1922

Entry

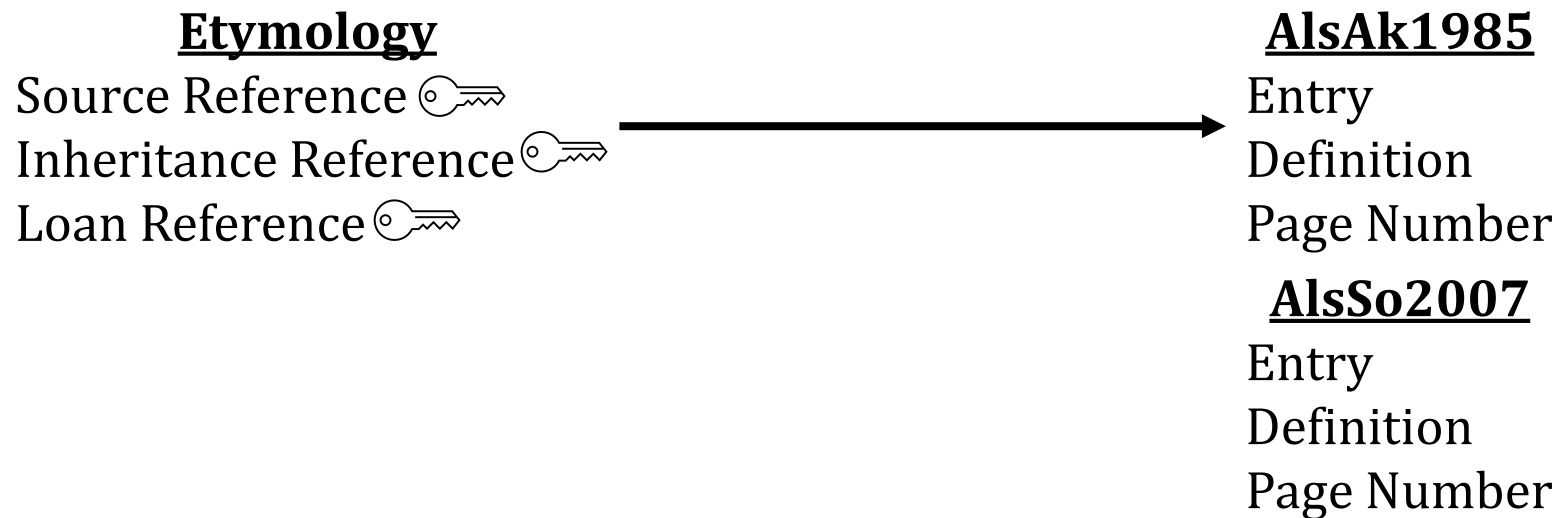
Definition

Page Number

123 page ▼	A-Z entry ▼	A-Z def-dutch ▼
187	kědih	langstaartige slankapen, witbuikige
209	sigöm	de gewone zwarte suikermier
187	imbau	de zwarte gibbon, of siamang

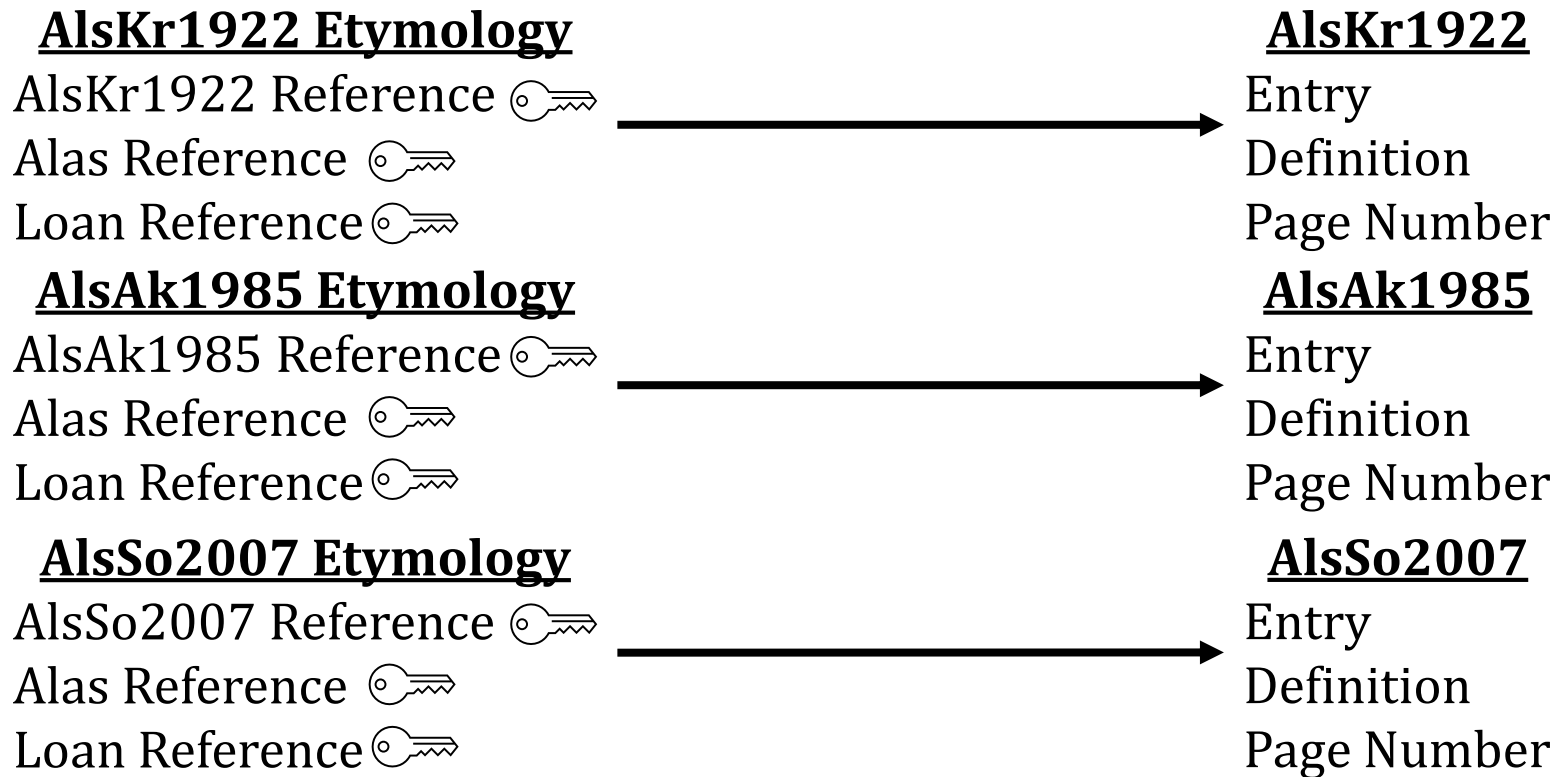
Historical linguistics data structuring

So, how *might* we structure data?



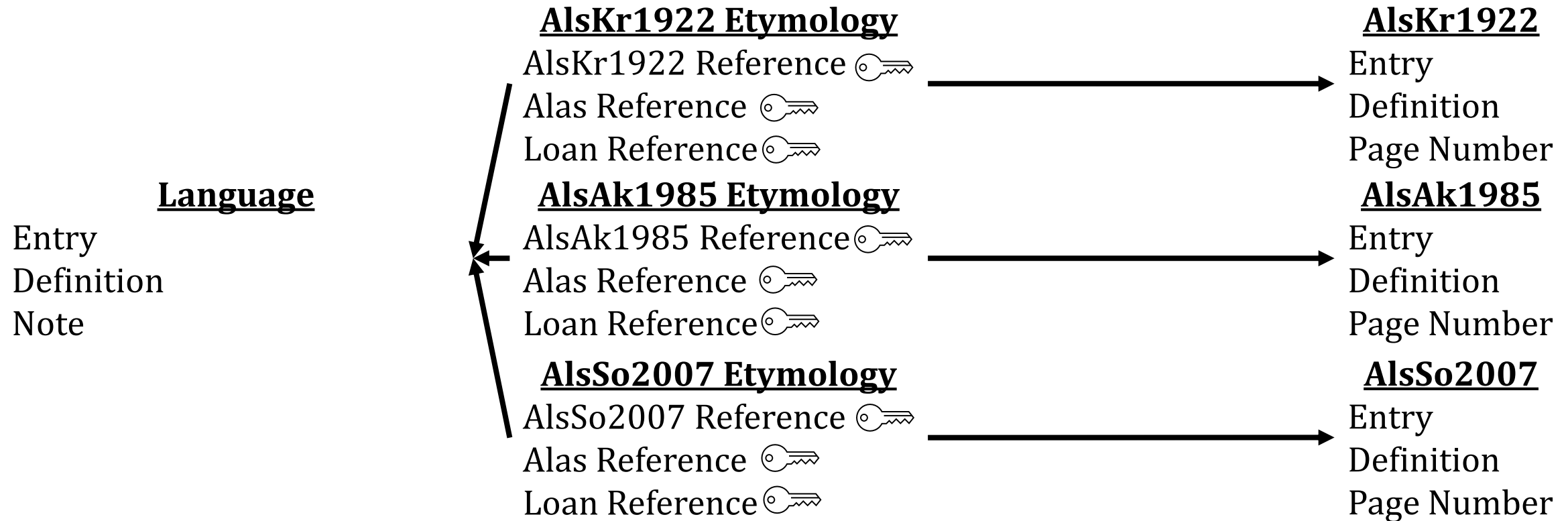
Historical linguistics data structuring

So, how *might* we structure data?



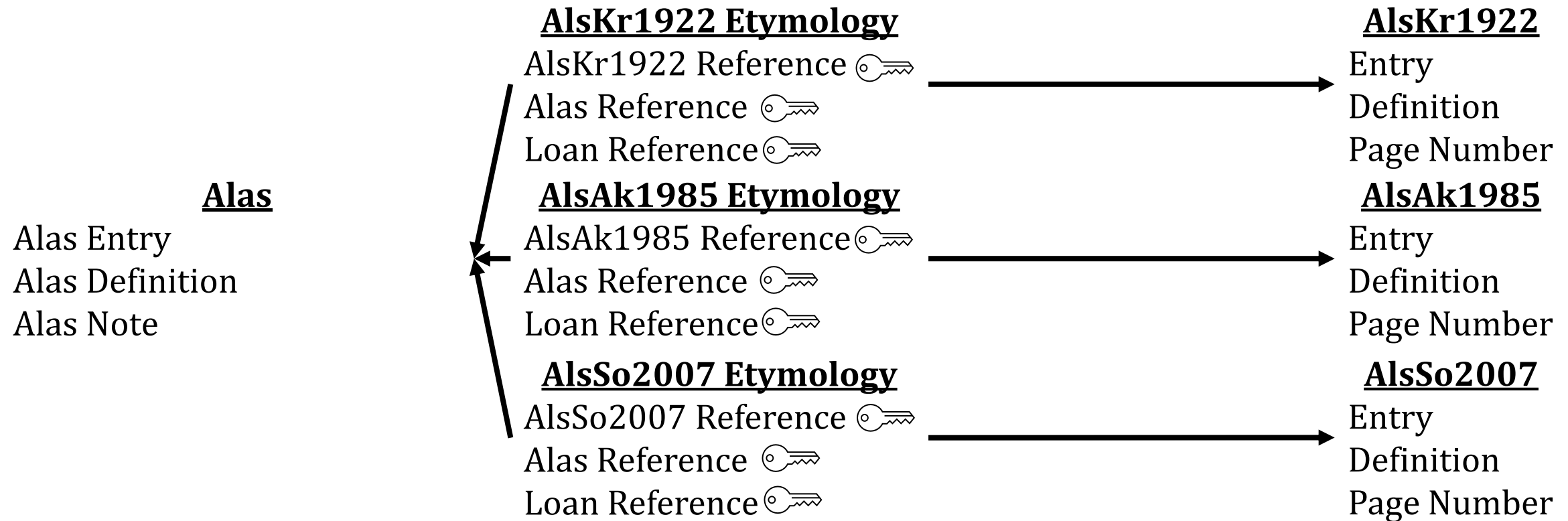
Historical linguistics data structuring

So, how *might* we structure data?



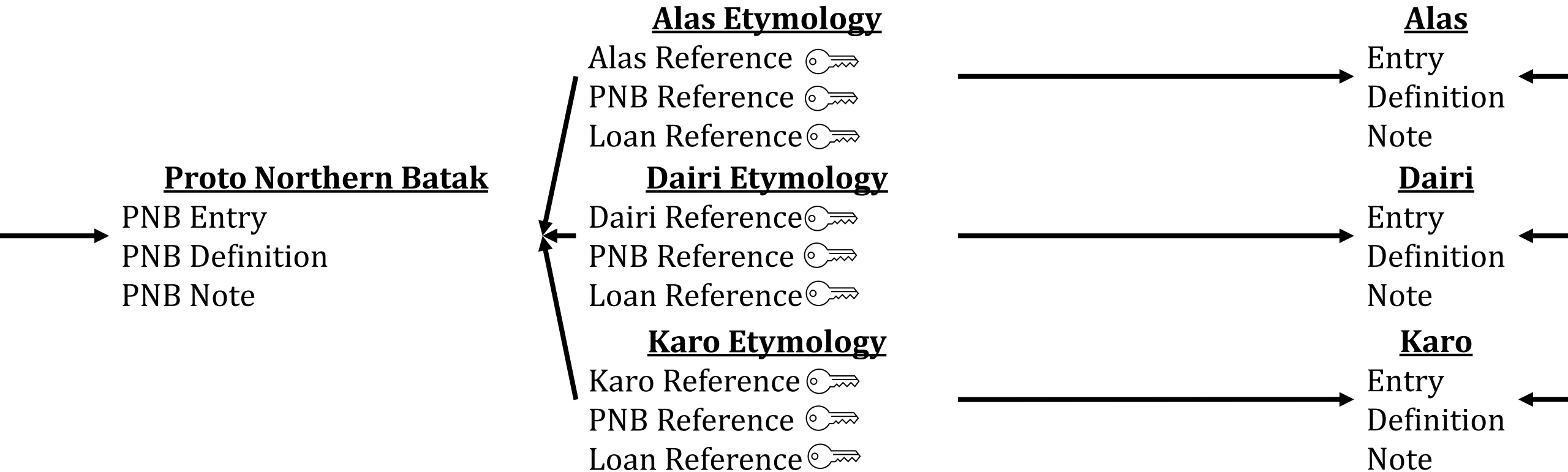
Historical linguistics data structuring

So, how *might* we structure data?



Historical linguistics data structuring

So, how *might* we structure data?



Part 4: Usability of database tools

Relational databases

What are databases?

- Includes *flat files*
- Includes *XML files*
- Includes *database management systems* (DBMSs)



database

/ˈdeɪtəbeɪs/

noun

plural noun: **databases**

a structured set of data held in a computer, especially one that is accessible in various ways.

DBMSs: Essentially fast, efficient, and secure storage of large, organized data

Relational DBMSs (RDBMSs):

- Organizational system akin to multiple sheets on a spreadsheet (but they need not all be open)
- RDBMSs are ubiquitous in nearly all other domains (Miedema et al. 2023: 16)

“Relational databases are ubiquitous and conceptually simple for all majors to understand based on the analogy of a spreadsheet with named columns.” — (Goelman & Dietrich 2018: 320)

Relational databases

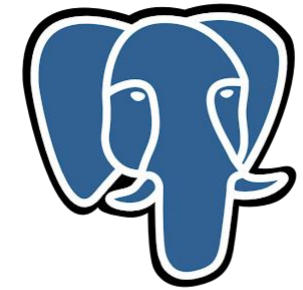
Many RDBMSs are *free, open-source, and extensively documented*

- MySQL, MongoDB, [PostgreSQL](#)

Many are available for *all platforms* (Windows, Mac, Linux)

They can be *edited directly* (like your favorite spreadsheet editor!)

- Datagrip (✓ open-source, ~ free – for students & faculty)
- DBVisualizer (✗ open-source, ✓ free)
- [DBeaver](#) (✓ open-source, ✓ free)



PostgreSQL



	A	B	C	D	E	
1	semdomnum ▼	semdomdesc ▼	hollenummer ▼	page ▼	entry ▼	def-dutch ▼
2	2.1.1	Face	3	32	bangòn	gezicht, aangezicht
3	2.1.1	Skull	5	32	lalar oeloe	schedel
4	2.1.5	Hair	6	32	bó oeloe	hoofdhaar
5	2.1.1.2	Ear	9	32	kójó kójó	oor
6	2.1.1.1	Eyelid	12	32	-	ooglid
7	2.1.5	Crown of Head	8	32	boefoes	kruin
8	2.1.1.1	Eyebrow	15	32	ling-ar mata	wenkbrauw
9	2.1.1.1	Eye	11	32	mata	oog
10	2.1.1.1	Eyelash	13	32	-	oogharen, wimpers
11	2.1.1.1	Eyebrow	14	32	-	wenkbrauwen
12	3.5.6.5	Tears	16	32	-	tranen
13	3.5.6.5	Tears	17	32	òwil mata	traan
14	2.1.1.3	Nostril	19	32	-	neusgat
15	2.2.4	Mucus	20	32	bèngòh bèngòh	snot
16	2.1.1.3	Nose	18	32	ichóng	neus
17	2.1.1.4	Mouth	23	32	-	uitwendige mond
18	2.1.1.4	Palate	32	32	teumbah teumbah	verhemelte
19	2.1.1	Cheek	21	32	asang	wang
20	2.1.1.4	Mouth	22	32	babah	mond
21	2.1.1.4	Mouth	24	32	-	inwendige mond
22	2.1.1.4	Lip	26	32	-	lip
23	2.1.1.4	Lip	25	32	bifil	lippen
24	2.1.8	Brain	36	32	-	hersenen
25	2.1.5	Mustache	27	32	koemi	knevel
26	2.1.5	Goatee	28	32	djanggoet	sik
27	2.1.1	Chin	29	32	seungeu	kin
28	2.1.5	Beard	30	32	baoe	baard
29	2 1 1 4	Tonone	31	32	dilah	tono

	 AZ semdomnum	AZ semdomdesc	AZ hollenummer	123 page	AZ entry	AZ def-dutch
1	2.1.1	Face	3	32	bangòn	gezicht, aangezicht
2	2.1.1	Skull	5	32	lalar oeloe	schedel
3	2.1.5	Hair	6	32	bó oeloe	hoofdhaar
4	2.1.1.2	Ear	9	32	kójó kójó	oor
5	2.1.1.1	Eyelid	12	32	-	ooglid
6	2.1.5	Crown of Head	8	32	boefoes	kruin
7	2.1.1.1	Eyebrow	15	32	ling-ar mata	wenkbrauw
8	2.1.1.1	Eye	11	32	mata	oog
9	2.1.1.1	Eyelash	13	32	-	oogharen, wimpers
10	2.1.1.1	Eyebrow	14	32	-	wenkbrauwen
11	3.5.6.5	Tears	16	32	-	tranen
12	3.5.6.5	Tears	17	32	òwil mata	traan
13	2.1.1.3	Nostril	19	32	-	neusgat
14	2.2.4	Mucus	20	32	bèngòh bèngòh	snot
15	2.1.1.3	Nose	18	32	ichóng	neus
16	2.1.1.4	Mouth	23	32	-	uitwendige mond
17	2.1.1.4	Palate	32	32	teumbah teumbah	verhemelte
18	2.1.1	Cheek	21	32	asang	wang
19	2.1.1.4	Mouth	22	32	babah	mond
20	2.1.1.4	Mouth	24	32	-	inwendige mond
21	2.1.1.4	Lip	26	32	-	lip

Database tools: Benefits

Easy to download and get started (**talk to me or email me!**)

Same old, familiar spreadsheet look and feel (with *more* and *better* functionality)

Form the basis for nearly all data-heavy web-based applications

Lots of advanced functionality, but nothing you need to learn to use it

Save database dumps, give version numbers, and upload for version control

All uses of a lexical item automatically linked to its citation if stored in the source file

Exceptionally fast for data access and editing

...

You can always **easily** back out! ($\Rightarrow$ convert it all to CSVs)

Part 5 (finally): Concluding thoughts

Conclusion

Final Notes: CLLD, CLDF, Other existing databases (Bower 2014, Greenhill 2015)

Assessing data organization methods early can assist greatly in analysis of historical data

Databases are accessible and familiar(*-ish*) format for organizing data

RDBMSs are the standard in all other domains

Database tables are easily exportable to other formats



Thank you!!!



Questions? Comments?

(Threats?)

References

- Akbar, Osra M., Syech Ahmaddin, & Hasanuddin Mardjo. 1985. *Kamus Alas-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Bowern, Claire. 2014. Data “big” and “small” – Examples from the Australian lexical database. *Linguistics Vanguard* 1(1), 295–303. doi:10.1515/lingvan-2014-1009
- Codd, E. F. 1990. *The relational model for database management*, version 2, 6. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.
- Greenhill, Simon J. 2015. TransNewGuinea.org: An online database of New Guinea languages. *PloS ONE* 10(10): e0141563. doi:10.1371/journal.pone.0141563
- Goelman, Don & Suzanne W. Dietrich. 2018. A visual introduction to conceptual database design for all. In Barnes, Tiffany, Daniel Garcia, Elizabeth K. Hawthorne, & Manuel A. Pérez Quiñones (eds.), *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 320–325. New York, NY: Association for Computing Machinery.
- Gwinn, J. M. 1992. Object-oriented programs in realtime. *ACM SIGPLAN Notices* 27(2), 47–56.
- J. Kreemer. 1922. *Atjèh: Algemeen samenvattend overzicht van land en volk van Atjèh en onderhoorigheden*, vol. 1. Leiden: E. J. Brill.
- Ku, Cyril S., Thomas J. Marlowe, & Nathan M. Mantell. 2006. Design patterns across software engineering and relational databases. In Dosch, Walter & William Perrizo (eds.), *Proceedings of the 15th International Conference on Software Engineering and Data Engineering*, 271–274.
- Miedema, Daphne, Michael Liut, George Fletcher, & Efthimia Aivaloglou. 2023. MSMI1: Towards a validated SQL misconceptions instrument. In Fisler, Kathi, Paul Denny, Diana Franklin, & Margaret Hamilton (eds.), *Proceedings of the 2023 ACM Conference on International Computing Education Research*, volume 2, 16–17. New York, NY: Association for Computing Machinery.
- Soravia, Giulio. 2007. *The Alas language (Northern Sumatra)*. Bologna: Università Bologna. (unpublished manuscript.)

References

Picture Credits:

Coupling & cohesion diagram: <https://medium.com/@granthopkins0315/beyond-the-monolith-embracing-flexibility-with-n-tiered-architectures-b87209528e9c>

Terrible text file: <https://www.withdata.com/blog/datafilesplitter/split-pipe-delimited-text-by-column-value.html>

Terrible excel spreadsheet 1: <https://x.com/BenCrellin/status/1511368354363289604?lang=ar>

Dilbert #1: <https://www.applicoinc.com/blog/spreadsheets-sabotaging-business/>

Dilbert #2: <https://www.auditexcel.co.za/blog/dilberts-understanding-of-spreadsheets/?srsltid=AfmBOop59JpTKmr350Ec0CzOhY4LsiUIj4L1NaHvkeZka9hrEID2GFvA>

PostgreSQL logo: <https://www.arbin.com/software-solution/introduction-postgresql.html>

DBeaver logo: <https://dvbrandt90.medium.com/dbeaver-just-build-the-dam-thing-76d710fed24e>