

语言学软件
ELAN & FLEX

1

陈德宁

2020-08-18

【语言学软件】这门课的目标

大体目标 “ELAN → FLEX → ELAN” 工作流程

第一节：ELAN



第二节：FLEX

第三节：ELAN->FLEX->ELAN



ELAN 简介

- 由马克斯普朗克心理语言学学院开发
- 注释软件，可以创造，编辑，并搜索音像数据的注释
- 主要功能：切分、处理自然的语言资料、建立语料库
- 创造有时间对齐的标注

下载：<https://tla.mpi.nl/tools/tla-tools/elan/>

版本：4.9

FLEX 简介



- 由SIL组织开发
- Shoebox > Toolbox > FLEX (2002-2020)
- 主要功能：语素分析的自动化

下载：<http://software.sil.org/fieldworks/download/fw-8312/>

“FieldWorks SE Full”

FLEX -



1 词 mǎ.ɿzìlè
语素 mǎ.ɿzì -lè
词汇注释 Chn 羌 复数
词汇注释 Eng Qiang PL

自由翻译 Eng Qiang people,
Chn 羌族人

2 词 dé jèwànàe # pàetèi
语素 dé jè -wà -nàe pǎ tèi
词汇注释 Chn 新年 到 将行体 链接 猪 杀
词汇注释 Eng newyears arrive PROS LNK pig kill

自由翻译 Eng when it is going to be newyears (we) kill the pigs.
Chn 要过年的时候杀猪

ELAN-FLEX-ELAN：传说中的“地狱三头犬”

难度大

软件经常出问题

复杂、枯燥、麻烦、等等



为什么要学ELAN-FLEX-
ELAN?

我们为什么要学这个工作流程？

外在原因

- 在濒危语言保护工作已经是一种标准
- 濒危语言保护项目会看好这些软件

我们为什么要学这个工作流程？

内在原因

- 语言分析的自动化

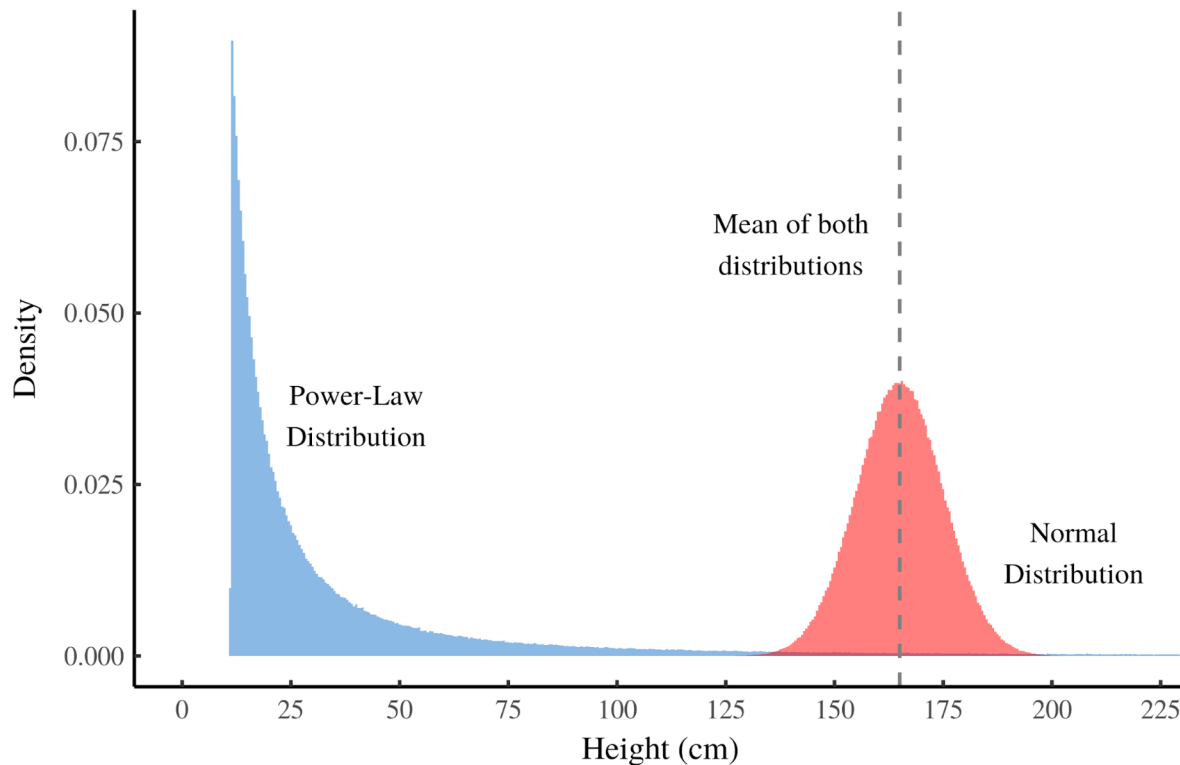
自动化的重要性：齐夫定律

是由哈佛大学的语言学家乔治·金斯利·齐夫（George Kingsley Zipf）于1949年发表的实验定律。

它可以表述为：在自然语言的语料库里，一个单词出现的频率与它在频率表里的排名成反比。

所以，频率最高的单词出现的频率大约是出现频率第二位的单词的2倍，而出现频率第二位的单词则是出现频率第四位的单词的2倍。

齐夫定律 vs. 常态分布



布朗语料库：100万字的自然语料库

在布朗语料库中，“the”、“of”、“and”是出现频率最前的三个单词。

大约占整个语料库100万个单词中的7%、3.6%、2.9%，

其比例约为 6 : 3 : 2。

满足齐夫定律中的描述。

仅仅前135个字汇就占了布朗语料库的一半！

本节课的目标

学会用ELAN、介绍ELAN的文件分类

导入媒体文件以及模板文件

使用ELAN的不同模式：标注模式、切分模式、等等

创建 .eaf 文件

分配作业

ELAN

处理语料的手段

手写



OUR STORY BEGINS

tə x' tze: p' wáj o k' tō huānāe l' z
k' tō tōw tsi-ɣiəmāe jōw p' k' p' k' tō.
nə ji-ɣ' hā-tō-wūnāe tō-g' hē g'
ɛj pi tō pu-nəe jōw k' tō: nē ɔ-k' tē
si-k' tē x' z nə tōw tō x' tsi-j' tō tō
* rest of horse tense.
ji-t' jāk' hē-ɣ' -zū-w' hēj tō tō tōw
huānāe ^{down} l' z h' en hō tō tōw
tō tō ^R ɣi-w' -ji nē | j' z l' k' j' nē 9
k' tō tōw ǎ! ũ k' i pi hē-ɣ' k' k' tō
nē t' ka pi nē
ũ ɔ-k' tōw hāmā s' ɛj d' z' m' tē ki
tōj ka d' z' p' k' l' z (ki) -notant
k' a d' z' p' k' l' z j' -j tōmāe hō tō
h' ɛj: k' a tō tō | ǎ. ũj m' x' w' ǎ! |
9/5 9/5 OBS NEG - 501

文本文件



LD-002.txt - Notepad

File Edit Format View Help

tjaq
kwætca zəʃkjɛptæ dzəs mækʃɔjlu
thəs mækʃɔj
dzəs mækʃɔjiɔ
phəq səjmi ʃən dzəs ɲwəj
thəs mækʃɔjiɔ
atɕu zəpə pʈswəlupi tsətɪnɔŋ
khwəpjɛ zæɕ
khwəpjɛ papa ʃən thəs ɲwəj
thala kəsauwan tsa
ej qa pəsin tsax
dzəs mækʃɔj
thəs mækʃətʃɔj
ndzəja zwə kʃɔjini
qa ndzəja zwə mimikɕitiji
kəj
kə kə kətuf
əw aɲnɔŋ ɛtɕi tɔpij
a qatʃə tɕakmanajitʃɔj
a əw zəp tsərka qa pəsin tsax
dzəs mækʃətʃɔj
thəs mækʃətʃɔj
dzəs phəq səjmi dzəs ɲwətʃɔj
thəs zəpə pʈswə lupi tsətɪnaʃtɕu khwəpjɛ papani thəs mækʃtʃɔɕ
ndzəja zwəkʃɔji
-- ndzəja zwə mimikɕitɕiɕ

Word文件



15. tɕie-tse tsɿ-kie-ni ʔũ tɕuan tɕia pəɽ wa ma ləz cy-wa
son-small this-CL-TOP 2PS zhuan jia do-PRS or < Ch.? book study-PRS

16. qa muɽu tɕuan tɕia pəɽ-wa mu-hua ləz cy-wa
1PS sky-day zhuan jia do-PRS:1PS sky-day book study-PRS:1PS

17. pəjæ tɕie-ba² tɕiou pəjæpu-ti.
pəjæ son-big jiu pəjælypu-???

18. pəjælypu-ni tɕəwu pei tɕia wə nəp sə-so ki tɕi wa:nəm
pəjælypu-TOP 3PS ??? shaman ?? three-CL go able-PRS ???

19. o: nəje , mi pa sɿ-tɕi mi pa kie na-ni
INT ??? NEG-?? temple NEG-??? CONJ-TOP

20. tɕəwu: o: mɛs ma-ɕa-ni, ɕawuzumi tsɿ-la fia-ɕa kue.
3PS INT ??? NEG-cut TOP sacred.grove this-CL DIR-cut ???

21. ɕawuzumi tsɿ fia-ɕa-k na
sacred.grove this DIR-cut-go CONJ

22. zi-ni-pei ə-tsɿ su pəjæ pei ə-thu tɕi na-ni
demon-black-? one-? ? pəjæ ??? DIR-bump able CONJ-TOP

ELAN 格式

媒体文件：

.MP4, .MPG, MPEG, .MOV, MPG4, .QT

.wav, .mp3

视频
录音

模板文件：

.etf “Elan Tier file”

ELAN 层次模板

.pfsx “ELAN Preferences”

ELAN 层次设定

输出文件：

.eaf “ELAN file” -

ELAN 文件

.Flextext

Flex 文件

四川省黑水县洛多乡石古子村



2016

