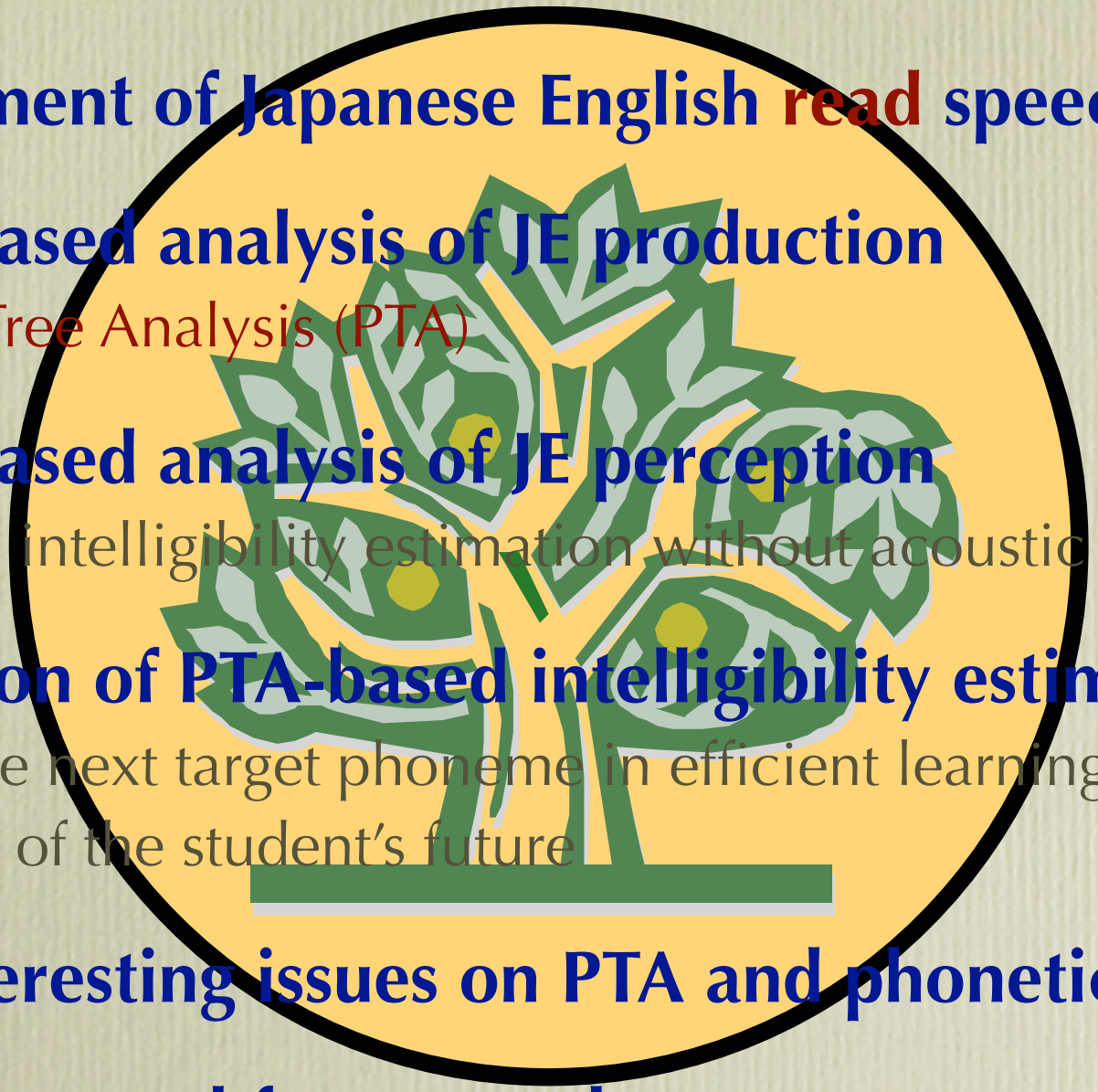
Phonetic Tree Analysis (PTA) and its application to estimation of the segmental intelligibility

Nobuaki MINEMATSU
UNIVERSITY OF TOKYO

Outline

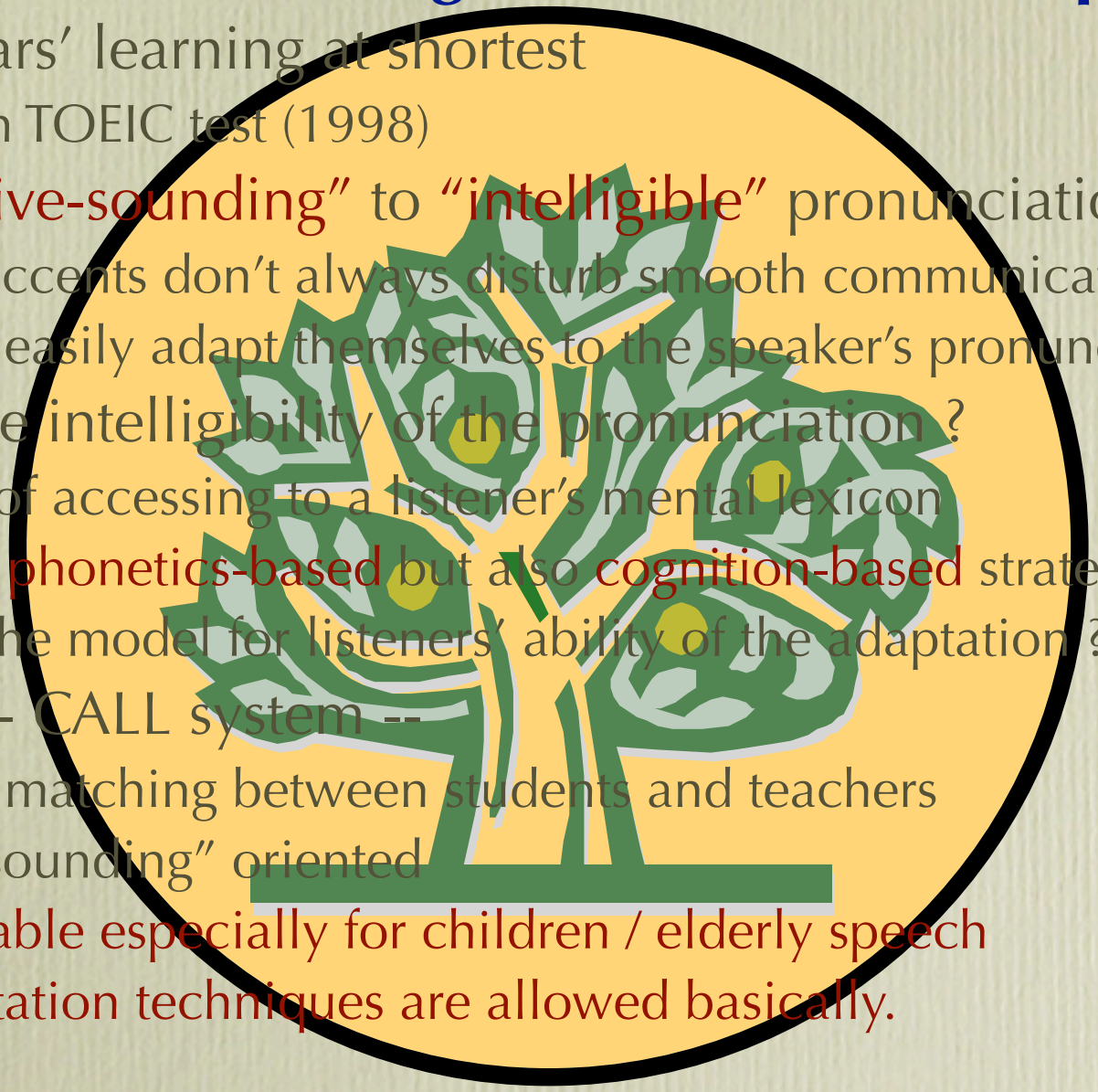
- Introduction
- Development of Japanese English read speech database
- Corpus-based analysis of JE production
  - Phonetic Tree Analysis (PTA)
- Corpus-based analysis of JE perception
  - Segmental intelligibility estimation without acoustic matching
- Application of PTA-based intelligibility estimation
  - What is the next target phoneme in efficient learning ?
  - Prediction of the student's future
- Some interesting issues on PTA and phonetics
- Conclusions and future works



Introduction (#1)

Current situation of English education in Japan

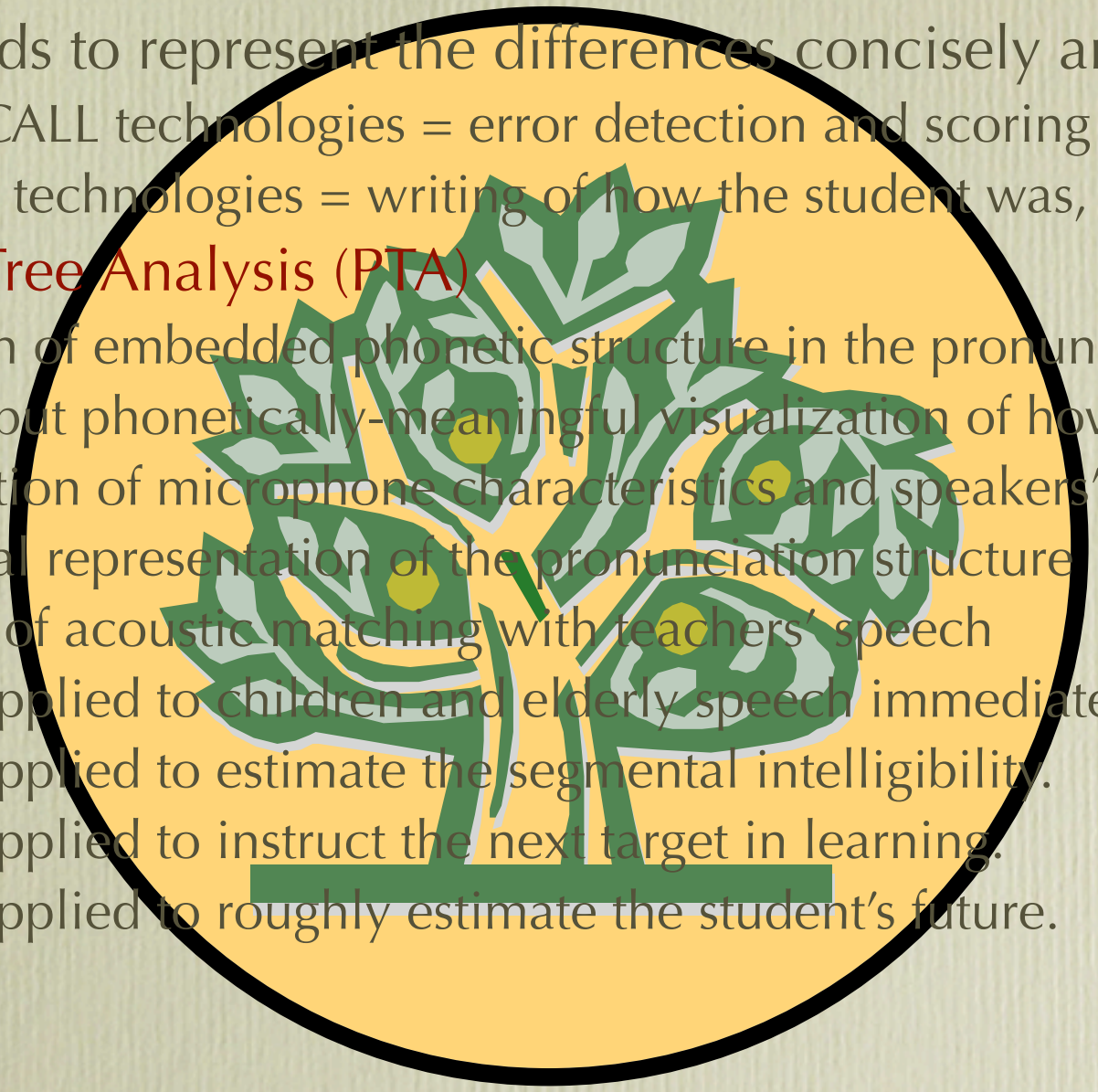
- 3+3+2 years' learning at shortest
  - 16 / 16 in TOEIC test (1998)
- From "native-sounding" to "intelligible" pronunciation
  - Foreign accents don't always disturb smooth communication.
  - Listeners easily adapt themselves to the speaker's pronunciation.
- What is the intelligibility of the pronunciation ?
  - Easiness of accessing to a listener's mental lexicon
  - Not only phonetics-based but also cognition-based strategy for LL
  - What is the model for listeners' ability of the adaptation ?
- A boom -- CALL system --
  - Acoustic matching between students and teachers
  - "Native-sounding" oriented
  - Still unstable especially for children / elderly speech
  - No adaptation techniques are allowed basically.



Introduction (#2)

No two students are the same.

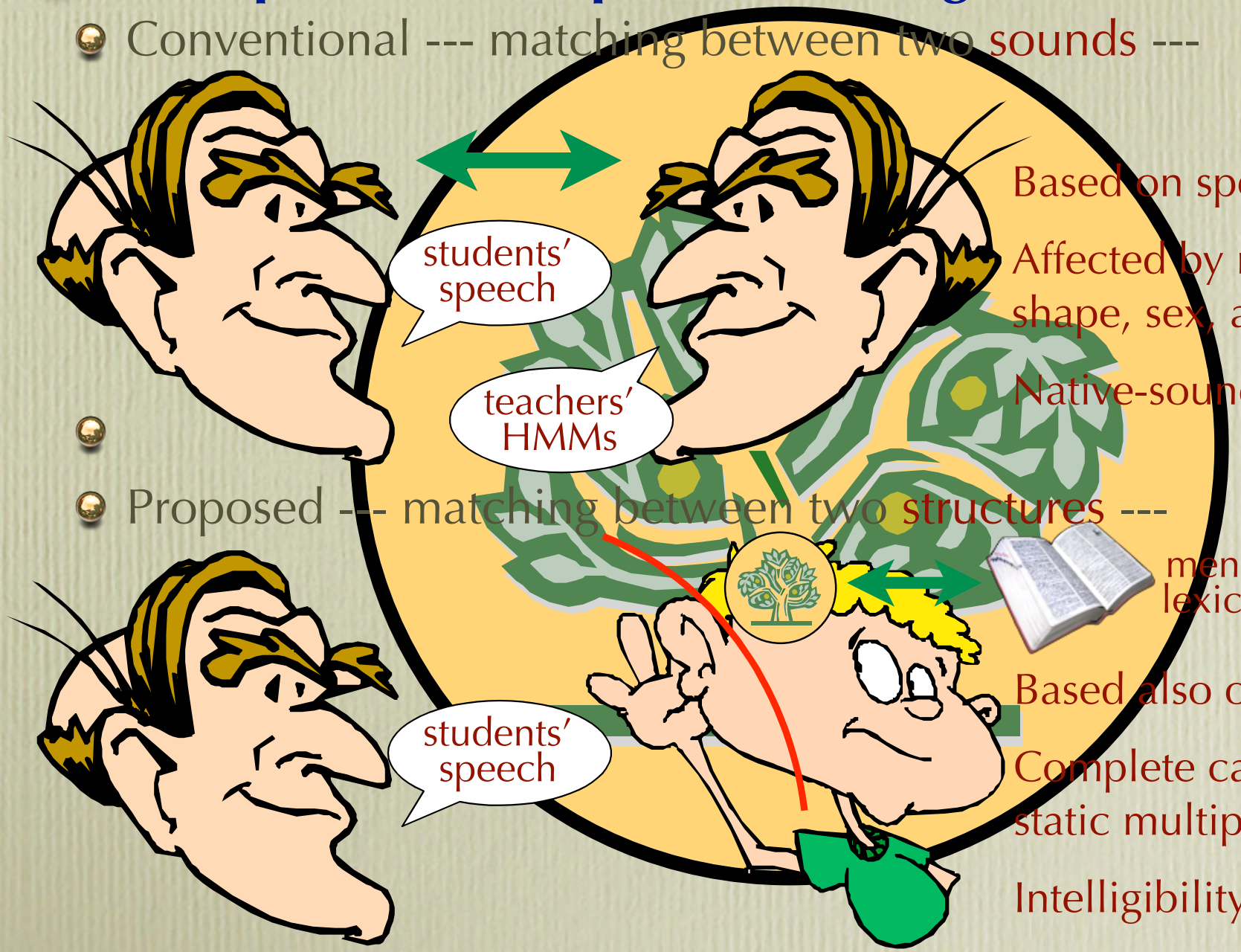
- No methods to represent the differences concisely and accurately
  - Current CALL technologies = error detection and scoring
  - Required technologies = writing of how the student was, is, and will be.
- Phonetic Tree Analysis (PTA)
  - Extraction of embedded phonetic structure in the pronunciation
  - Abstract but phonetically-meaningful visualization of how the student is
  - Cancellation of microphone characteristics and speakers' individuality
  - Perceptual representation of the pronunciation structure
  - No need of acoustic matching with teachers' speech
  - Can be applied to children and elderly speech immediately.
  - Can be applied to estimate the segmental intelligibility.
  - Can be applied to instruct the next target in learning.
  - Can be applied to roughly estimate the student's future.
  - :



Introduction (#3)

From phonetics to phonetics+cognition

- Conventional --- matching between two sounds ---



Based on speech sci. and eng.

Affected by mic, BN, size, shape, sex, age, individuality

Native-sounding oriented

- Proposed --- matching between two structures ---

Based also on cognition

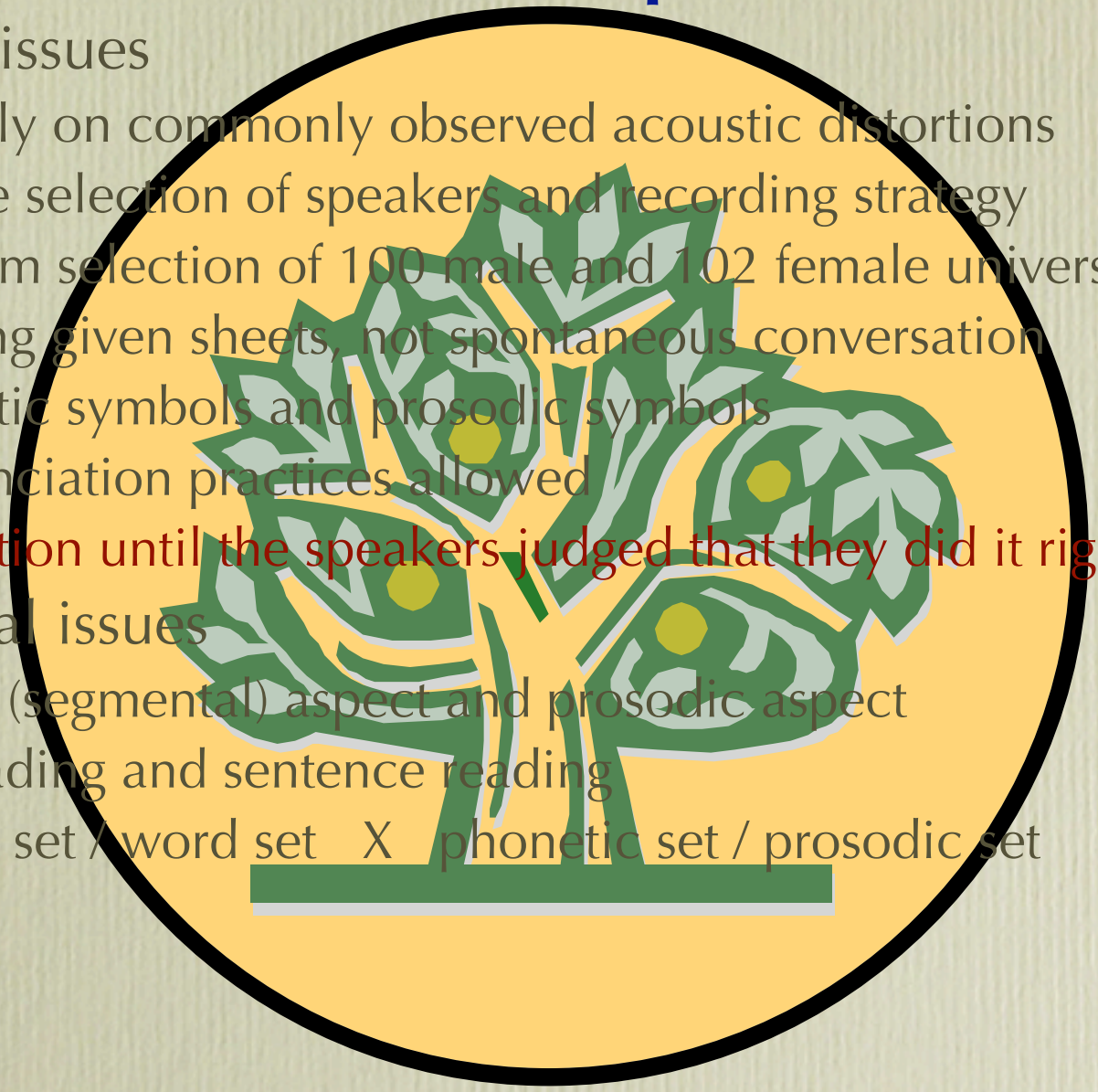
Complete cancellation of static multiplicative noise

Intelligibility oriented

Development of JE speech database (#1)

Technical and educational requirements for the design

- Technical issues
  - Focus only on commonly observed acoustic distortions
  - Adequate selection of speakers and recording strategy
    - Random selection of 100 male and 102 female university students
    - Reading given sheets, not spontaneous conversation
    - Phonetic symbols and prosodic symbols
    - Pronunciation practices allowed
    - Repetition until the speakers judged that they did it right.
- Educational issues
  - Phonetic (segmental) aspect and prosodic aspect
  - Word reading and sentence reading
  - Sentence set / word set X phonetic set / prosodic set



Development of JE speech database (#2)

Word / sentence sets for the phonetic aspect

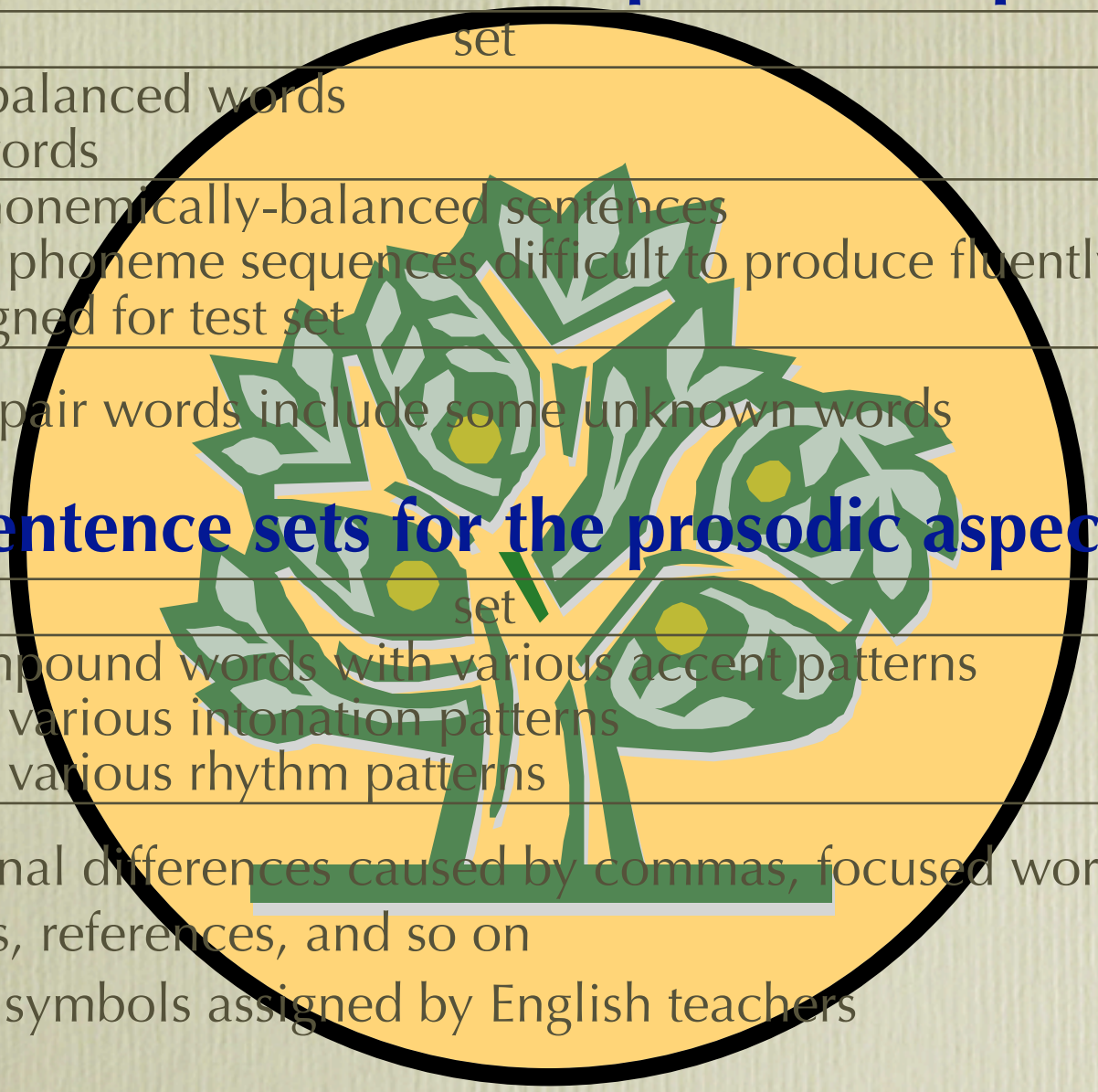
set	size
Phonemically-balanced words	300
Minimal pair words	600
TIMIT-based phonemically-balanced sentences	460
Sentences with phoneme sequences difficult to produce fluently	32
Sentences designed for test set	100

- Minimal pair words include some unknown words

Word / sentence sets for the prosodic aspect

set	size
Words and compound words with various accent patterns	109
Sentences with various intonation patterns	94
Sentences with various rhythm patterns	121

- Intonational differences caused by commas, focused words, syntactic structures, references, and so on
- Prosodic symbols assigned by English teachers



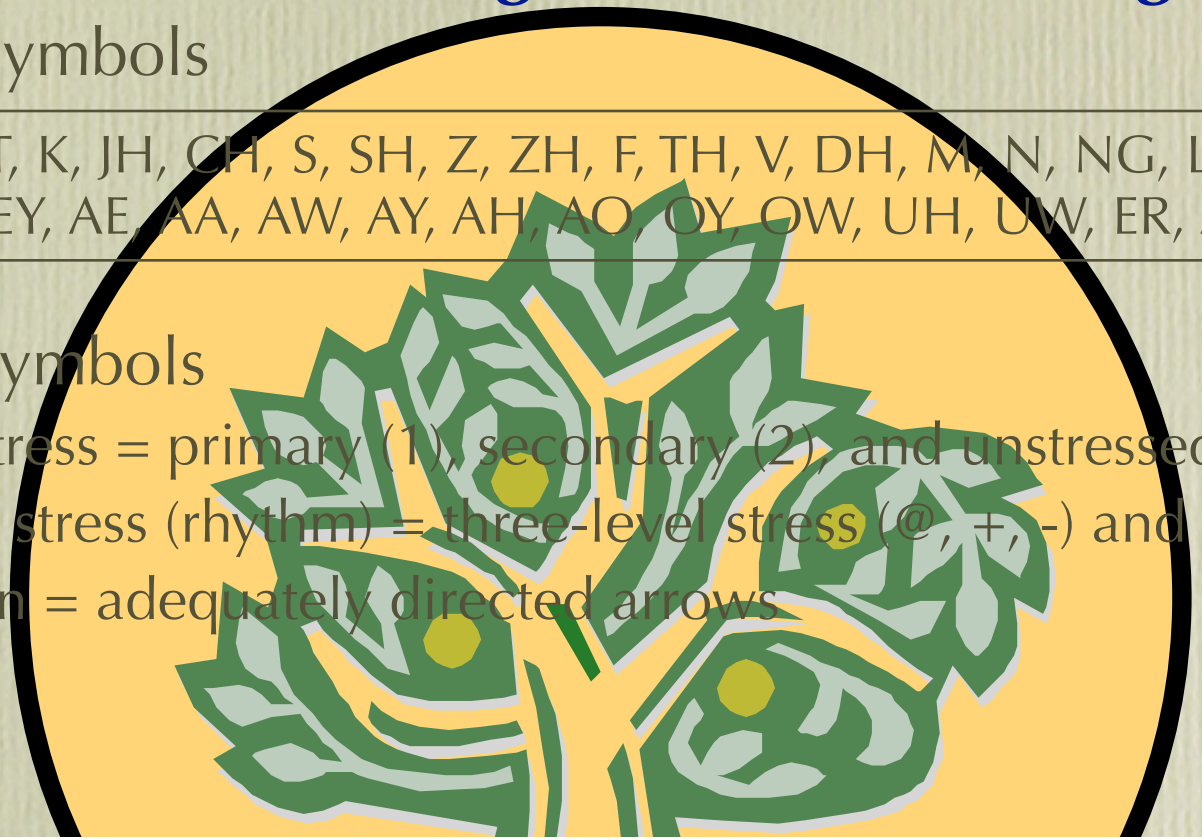
Development of JE speech database (#3)

Symbols and marks assigned to the reading sheets

- Phonetic symbols

B, D, G, P, T, K, JH, CH, S, SH, Z, ZH, F, TH, V, DH, M, N, NG, L, R, W, Y, HH,
IY, IH, EH, EY, AE, AA, AW, AY, AH, AO, OY, OW, UH, UW, ER, AXR, AX

- Prosodic symbols
  - Lexical stress = primary (1), secondary (2), and unstressed (0)
  - Sentence stress (rhythm) = three-level stress (@, +, -) and phrase break (/)
  - Intonation = adequately directed arrows

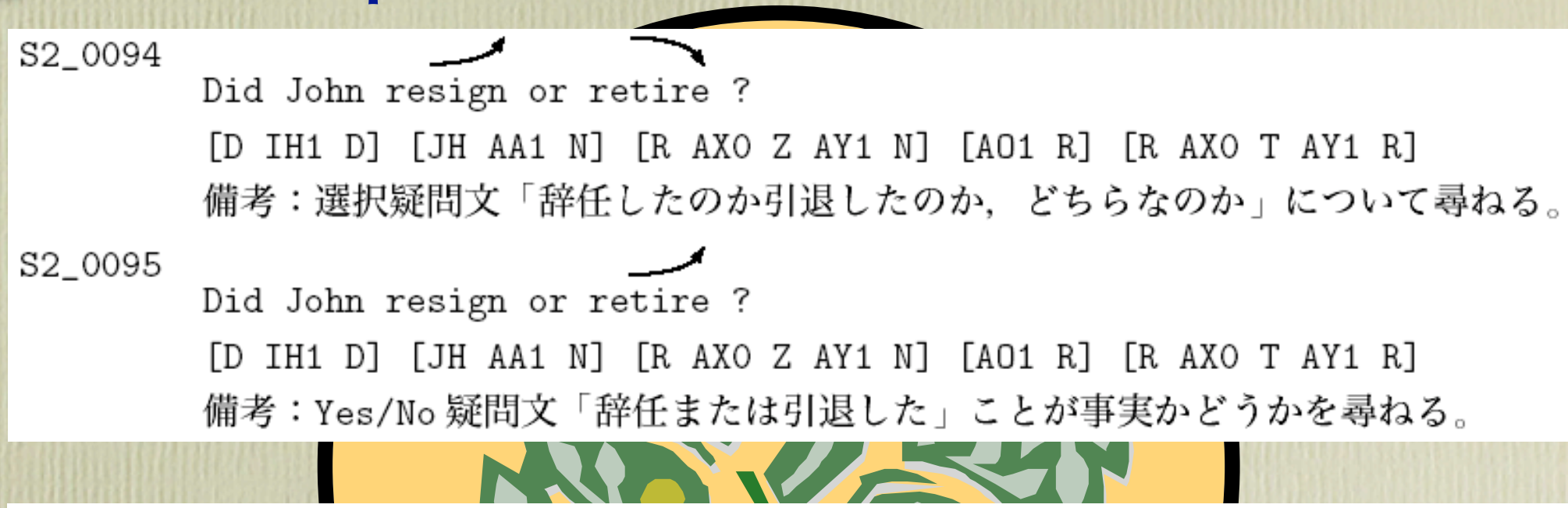


Examples

```
S1_0001  This was easy for us.
         [DH IH1 S] [W AA1 Z] [IY1 Z IYO] [F AO1 R] [AH1 S]
S1_0002  Is this seesaw safe ?
         [IH1 Z] [DH IH1 S] [S IY1 S AO2] [S EY1 F]
S1_0003  Those thieves stole thirty jewels.
         [DH OW1 Z] [TH IY1 V Z] [S T OW1 L] [TH ER1 T IYO] [JH UW1 AXO L Z]
```

Development of JE speech database (#4)

More examples



S2_0094
Did John resign or retire ?
[D IH1 D] [JH AA1 N] [R AXO Z AY1 N] [AO1 R] [R AXO T AY1 R]
備考：選択疑問文「辞任したのか引退したのか，どちらなのか」について尋ねる。

S2_0095
Did John resign or retire ?
[D IH1 D] [JH AA1 N] [R AXO Z AY1 N] [AO1 R] [R AXO T AY1 R]
備考：Yes/No 疑問文「辞任または引退した」ことが事実かどうかを尋ねる。

S1_0105 Come to tea.
/ + - @ /
[K AH1 M] [T UW1] [T IY1]

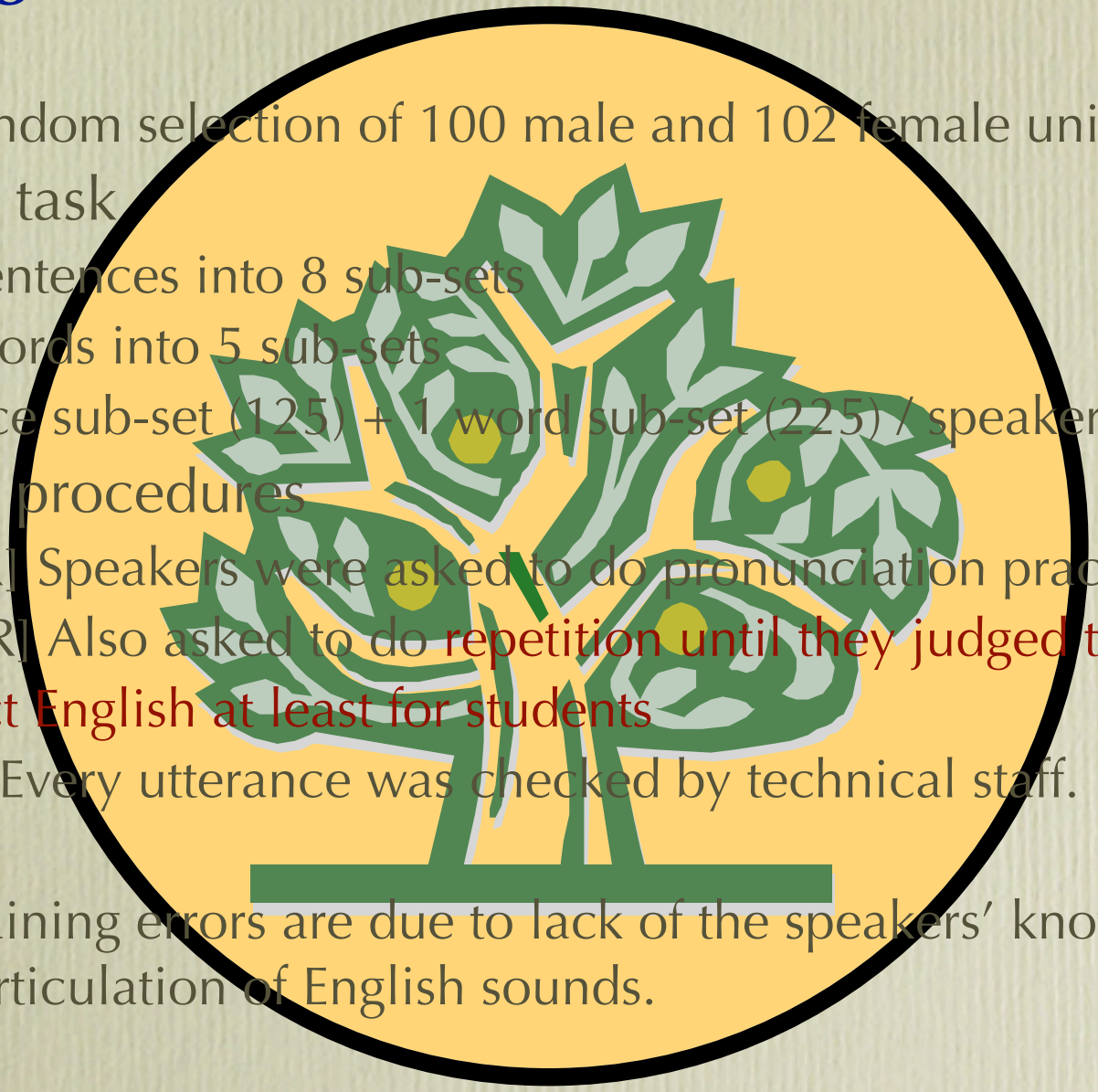
S1_0106 Come to tea with John.
/ + - + - @ /
[K AH1 M] [T UW1] [T IY1] [W IH1 DH] [JH AA1 N]

S1_0107 Come to tea with John and Mary.
/ + - @ / - + - @ -/
[K AH1 M] [T UW1] [T IY1] [W IH1 DH] [JH AA1 N] [AE1 N D] [M EH1 R IYO]

Development of JE speech database (#5)

Recording

- Speakers
  - Quasi-random selection of 100 male and 102 female univ. students
- Recording task
  - All the sentences into 8 sub-sets
  - All the words into 5 sub-sets
  - 1 sentence sub-set (125) + 1 word sub-set (225) / speaker
- Recording procedures
  - [Before R] Speakers were asked to do pronunciation practices
  - [During R] Also asked to do repetition until they judged they did it right.
    - Correct English at least for students
  - [After R] Every utterance was checked by technical staff.

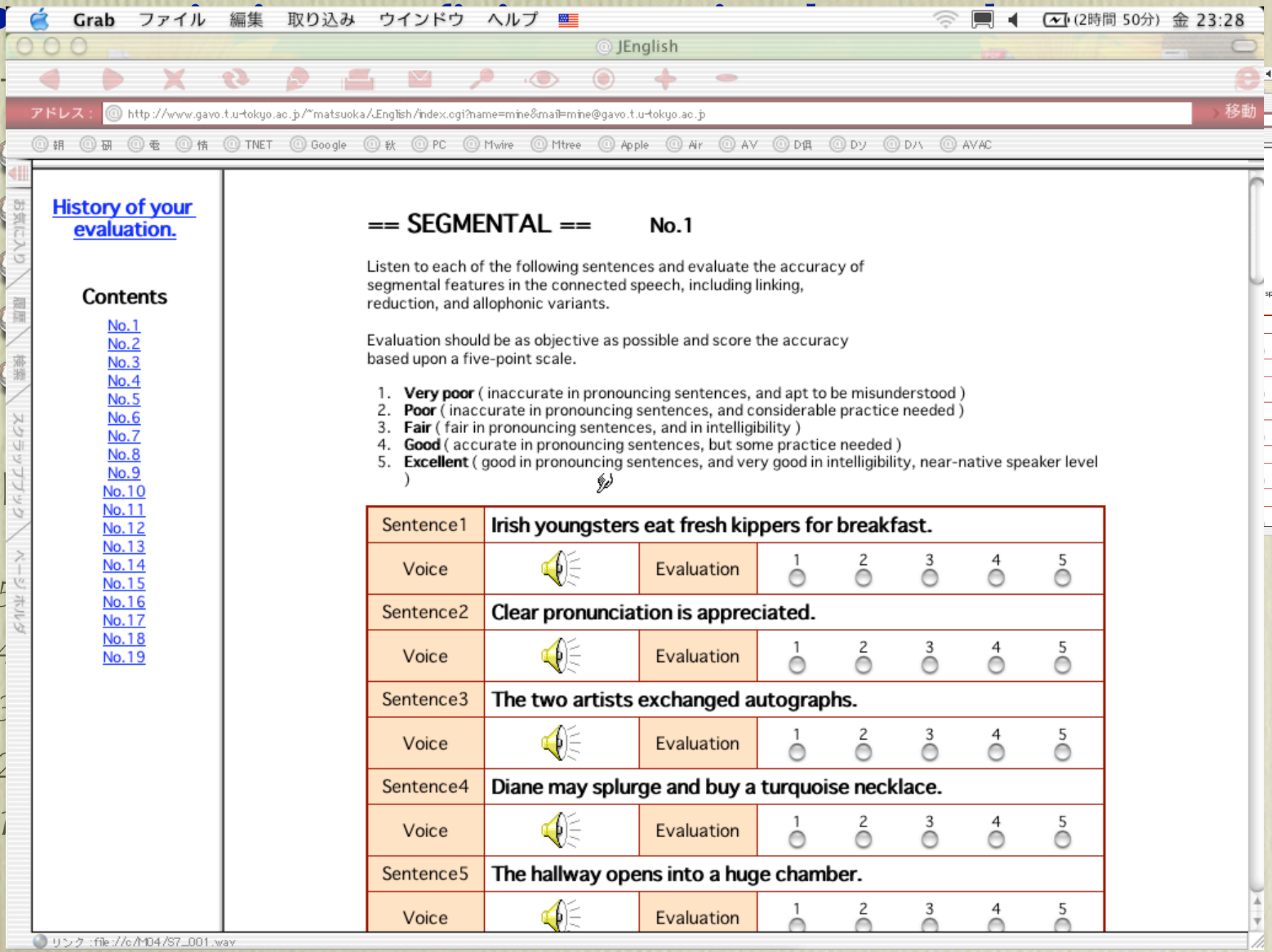


  - The remaining errors are due to lack of the speakers' knowledge of correct articulation of English sounds.

Development of JE speech database (#6)

P

#students

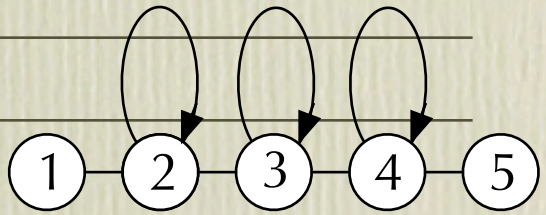


Corpus analysis of JE production (SI, #1)

Training of AE and JE SI-HMMs

- Monophones with a single mixture for easy visualization
- Transcription automatically generated with PRONLEX
  - Pronunciation errors were not represented in the transcription.

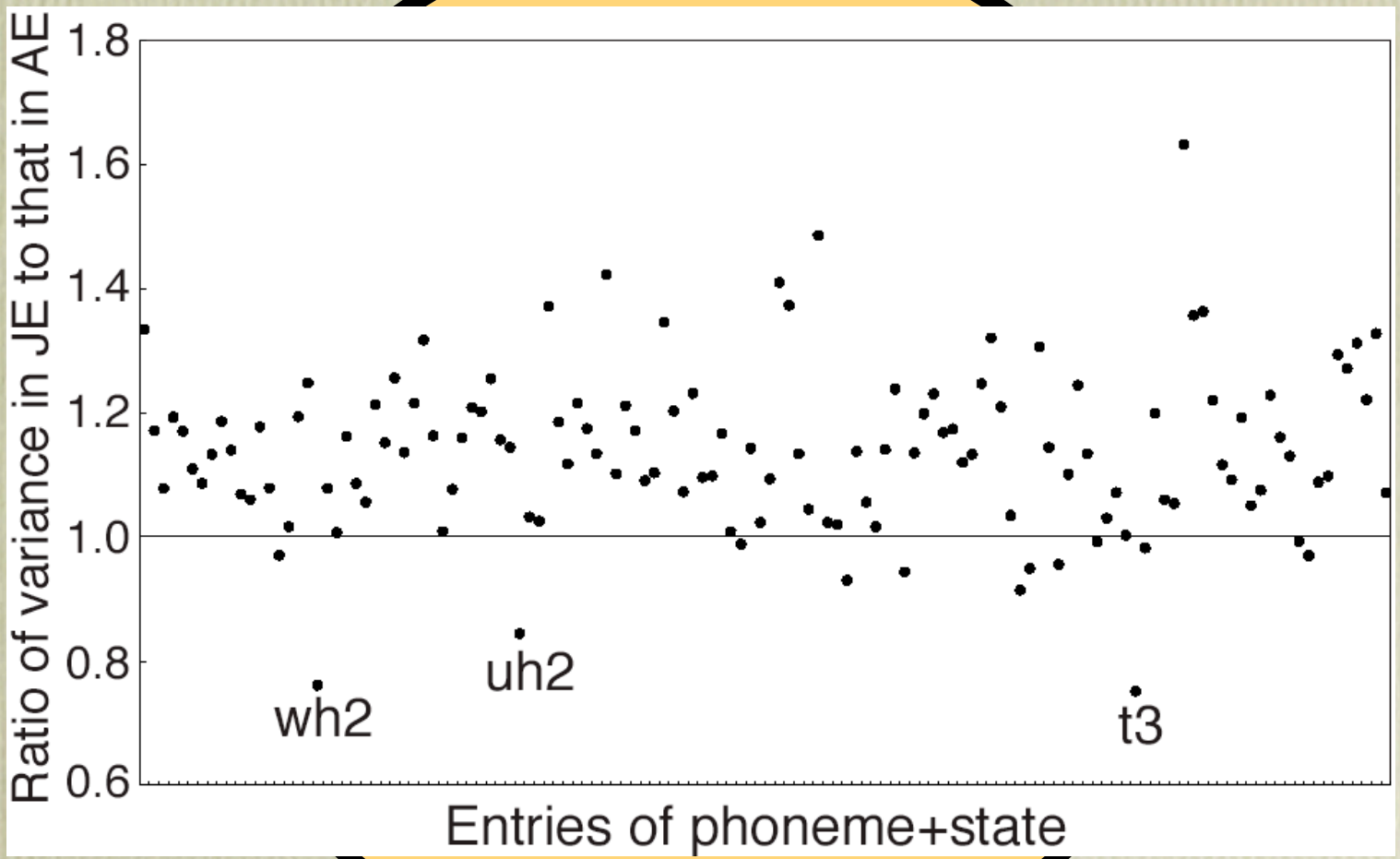
AD	16bit / 16kHz sampling
Window	Hamming, 25 ms length, 10 ms rate
Pre-emphasis	1-0.97 z-1
Parameters	12MFCC + 12ΔMFCC + ΔPower (25dim)
Phoneme set	ae, ah, ch, dh, eh, nx, wh, ih, jh, oy, er, sh, th, uh, aw, ay, zh, aa, b, ao, d, ey, f, g, hh, iy, k, l, m, o, ow, p, r, s, t, uw, v, w, ax, y, z
HMMs	1-mixture monophones with diagonal matrices (5st. + 3dis.)
Training	AE = 245 male Americans, 25,652 sentences (WSJ) JE = 68 male Japanese, 8,282 utterances
Initial models	Models built with TIMIT (4,346 utterances)



Corpus analysis of JE production (SI, #2)

Magnitude of variances of AE and JE

- Relative difference in averaged variances over cep. dimensions

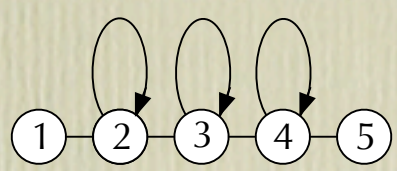


- JE > AE although #JE <<< #AE and JE are carefully read samples.
- Due to inter-speaker differences in pronunciation proficiency

Corpus analysis of JE production (SI, #3)

Phoneme pairs difficult for Japanese to discriminate

- Ratio of state distance in JE to that in AE



pair	s2	s3	s4	pair	s2	s3	s4
/r/&/l/	0.18	0.18	0.10	/hh/&/f/	0.31	0.27	0.58
/s/&/th/	0.09	0.03	0.10	/b/&/v/	0.94	0.79	0.40
/s/&/sh/	0.28	0.34	0.55	/ih/&/iy/	0.22	0.20	0.15
/th/&/sh/	0.23	0.32	0.49	/ih/&/y/	0.17	0.29	0.62
/z/&/zh/	0.36	0.49	0.76	/uh/&/uw/	0.26	0.23	0.30
/z/&/dh/	0.20	0.24	0.35	/ae/&/aa/	0.24	0.28	0.62
/z/&/jh/	0.21	0.37	0.62	/ae/&/ah/	[illegible]	0.17	0.12
/zh/&/jh/	0.31	0.32	0.56	/aa/&/ah/	[illegible]	0.13	0.46
/zh/&/dh/	0.27	0.31	0.38	/er/&/ah/	[illegible]	[illegible]	0.16
/dh/&/jh/	0.19	0.20	0.44	/er/&/aa/	[illegible]	[illegible]	0.26
/n/&/ng/	0.74	0.59	0.50	/er/&/ae/	0.15	0.09	0.22

- Larger confusion is clearly seen in each pair.
- Mid and low vowels are replaced by a Japanese mid and low vowel /あ/.

Corpus analysis of JE production (SI, #4)

Schwa and other vowels in AE and JE

- Five nearest phonemes (not vowels) to schwa in AE and JE

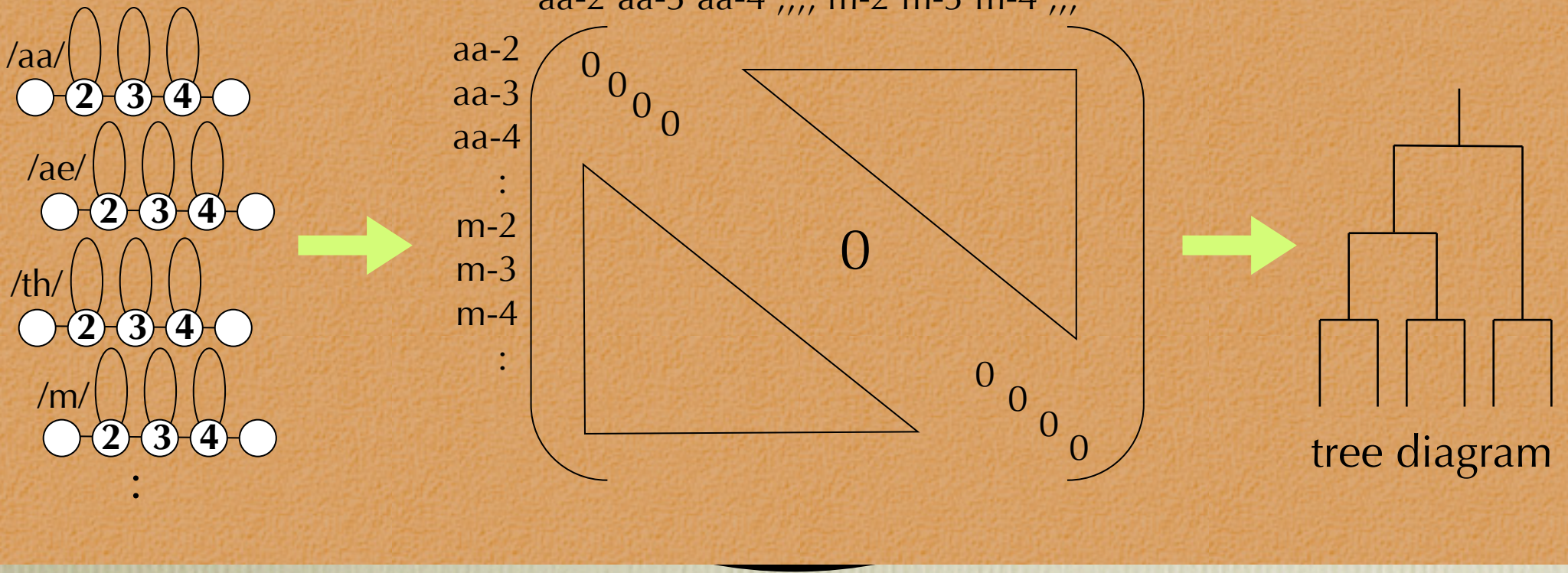
state	1st	2nd	3rd	4th	5th
ax2/AE	ih2(0.68)	uh2(0.73)	d4(0.75)	ah2(0.76)	eh2(0.86)
ax3/AE	ih3(0.87)	uh3(0.88)	eh4(0.93)	ae4(0.94)	uw4(0.96)
ax4/AE	uw4(0.69)	ih4(0.72)	uh4(0.76)	ah4(0.80)	eh4(0.84)
ax2/JE	ae2(0.46)	ah2(0.51)	aa2(0.51)	ay2(0.65)	aw2(0.69)
ax3/JE	ah3(0.57)	ae3(0.61)	aa3(0.72)	aw3(0.80)	uh3(0.87)
ax4/JE	ah4(0.54)	ae4(0.61)	aa4(0.73)	aw4(0.78)	uh4(0.86)

- Schwa is one of the most difficult sounds for Japanese to produce.
- Various vowels are found in AE but only mid and low vowels in JE.
- Mid and low vowels of JE = /あ/
  - Japanese perceive /あ/ in native schwa sounds.
  - Japanese produce /あ/ for schwa.

Corpus analysis of JE production (SI, #5)

Phonetic Tree Analysis (PTA) with SI HMMs

- State-level distance matrices for AE and JE
- Bhattacharyya distance measure
- Hierarchical clustering for each matrix with Ward's method



Corpus analysis of JE production (SI, #6)

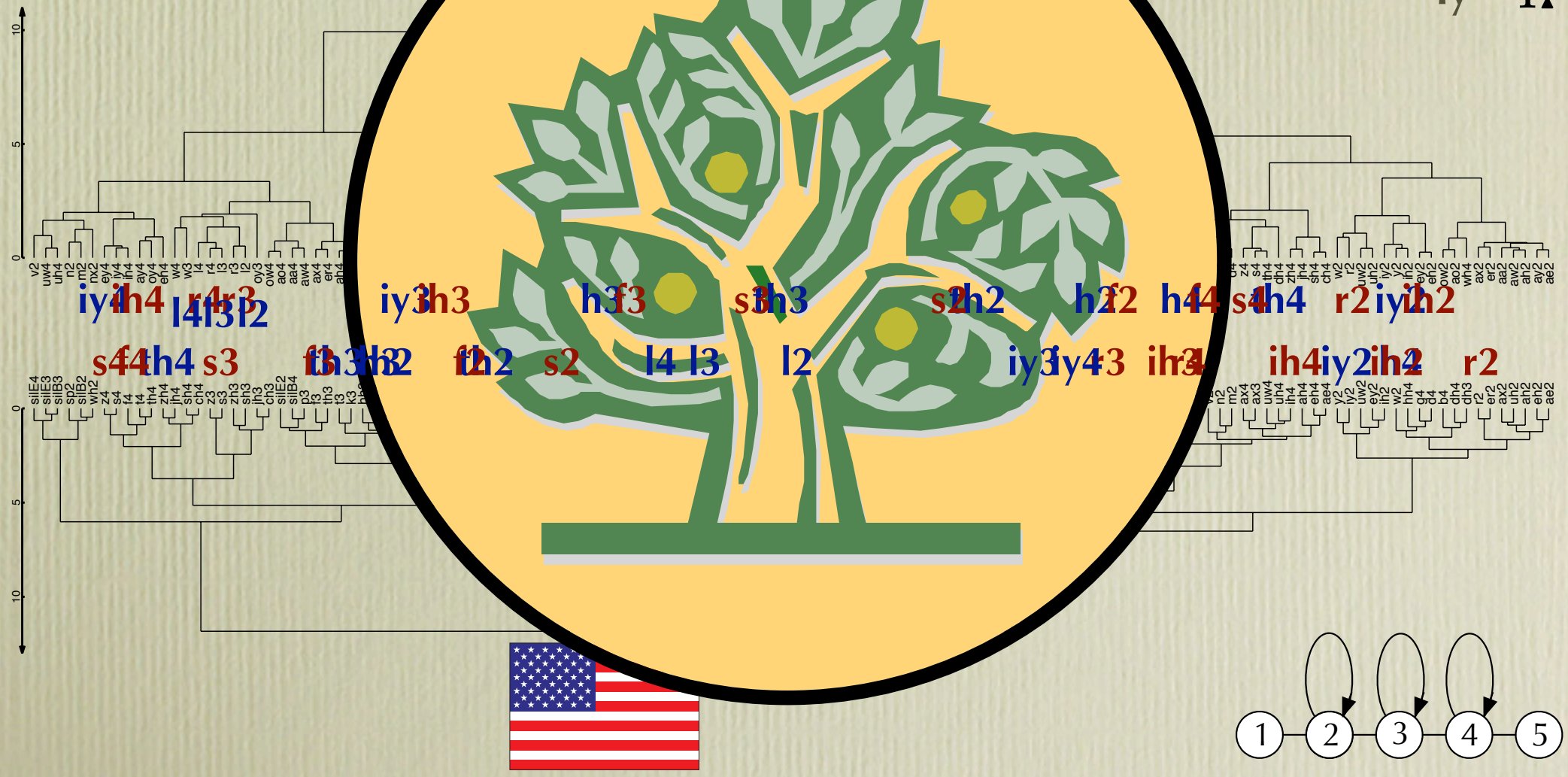
Where are the confusing phonemes in the two trees ?

r/l ? s/th ? f/h ? ih/iy ?

th= θ

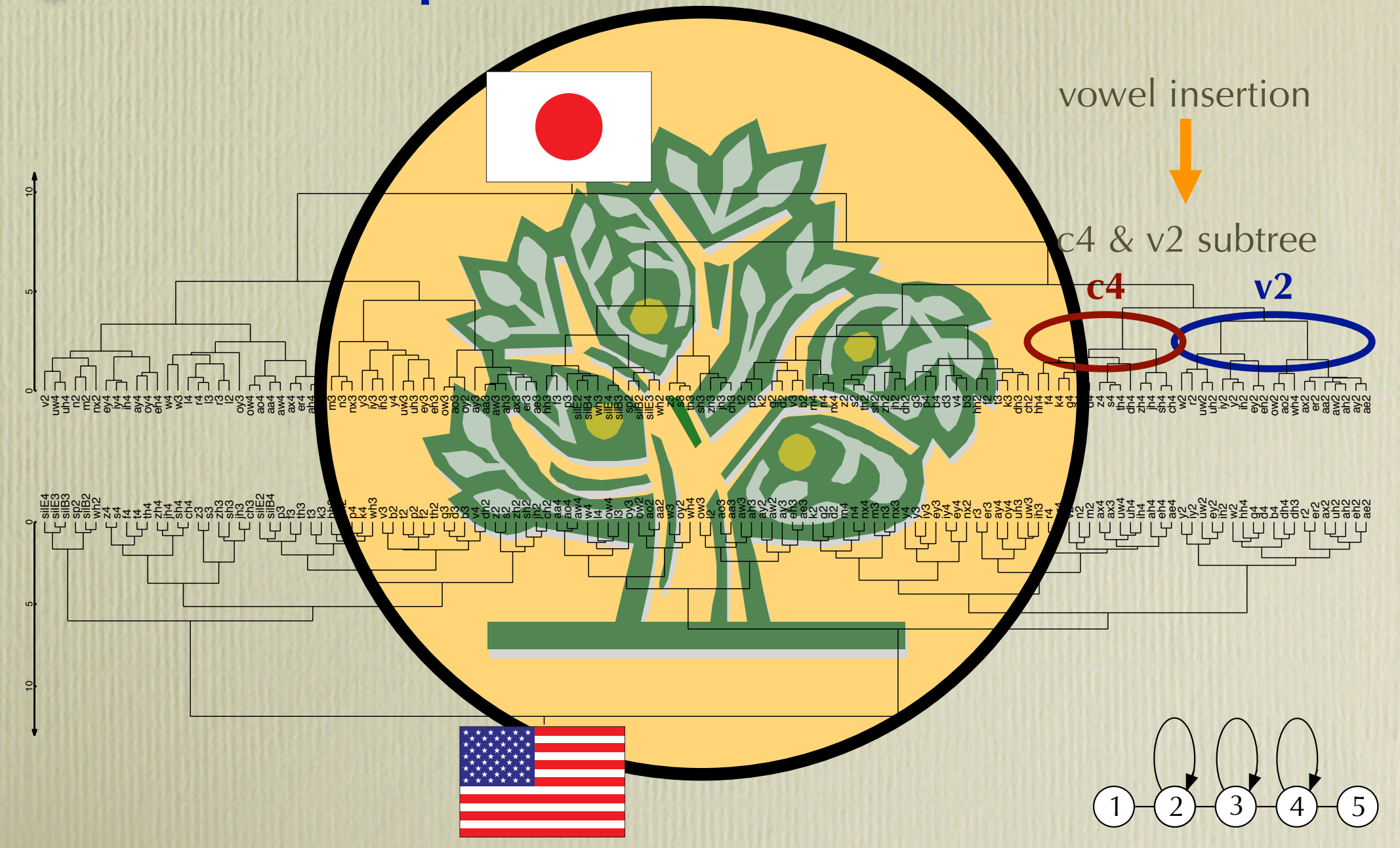
ih= ɪ

iy= iː



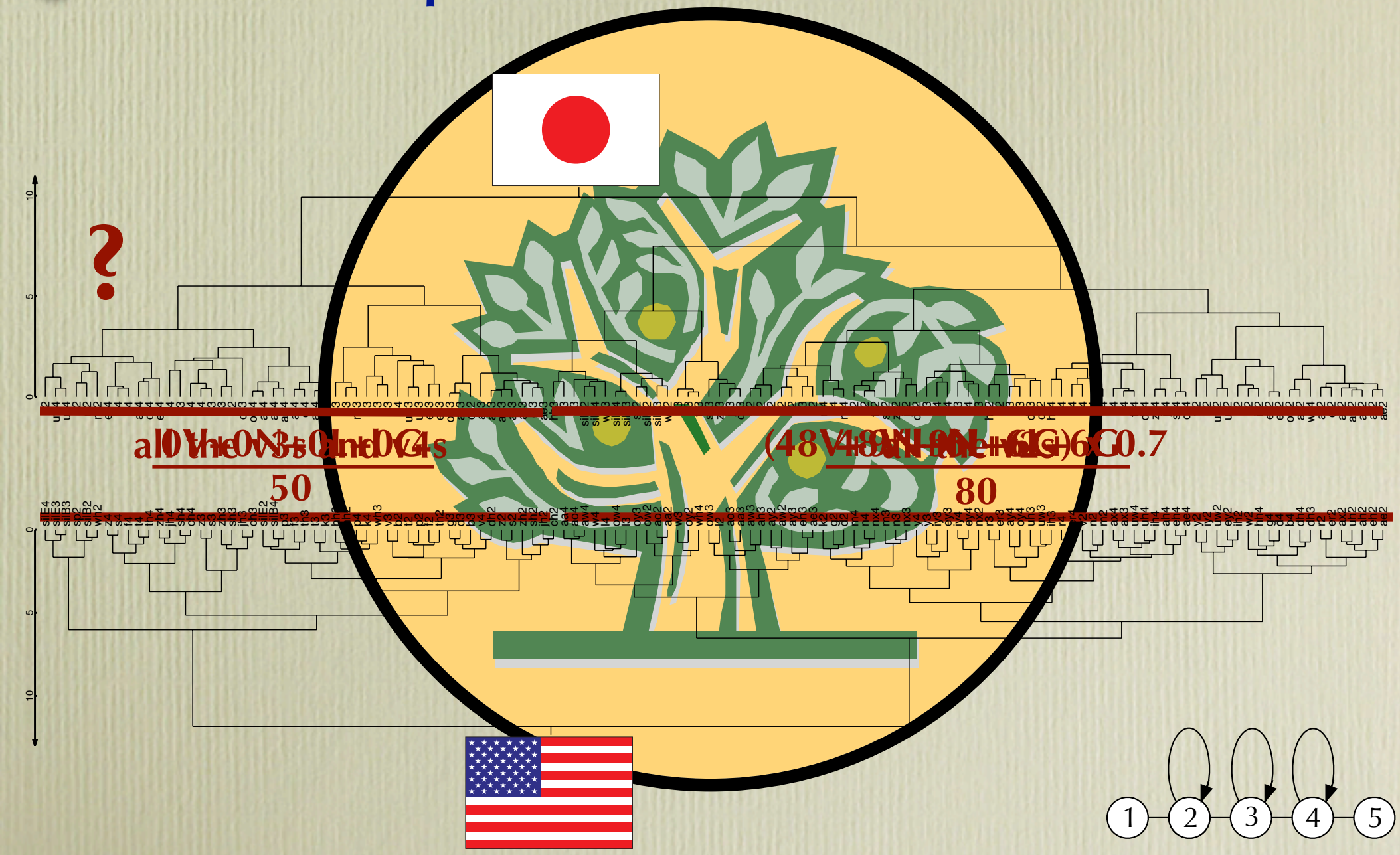
Corpus analysis of JE production (SI, #7)

Phonetic interpretation of the tree structure



Corpus analysis of JE production (SI, #8)

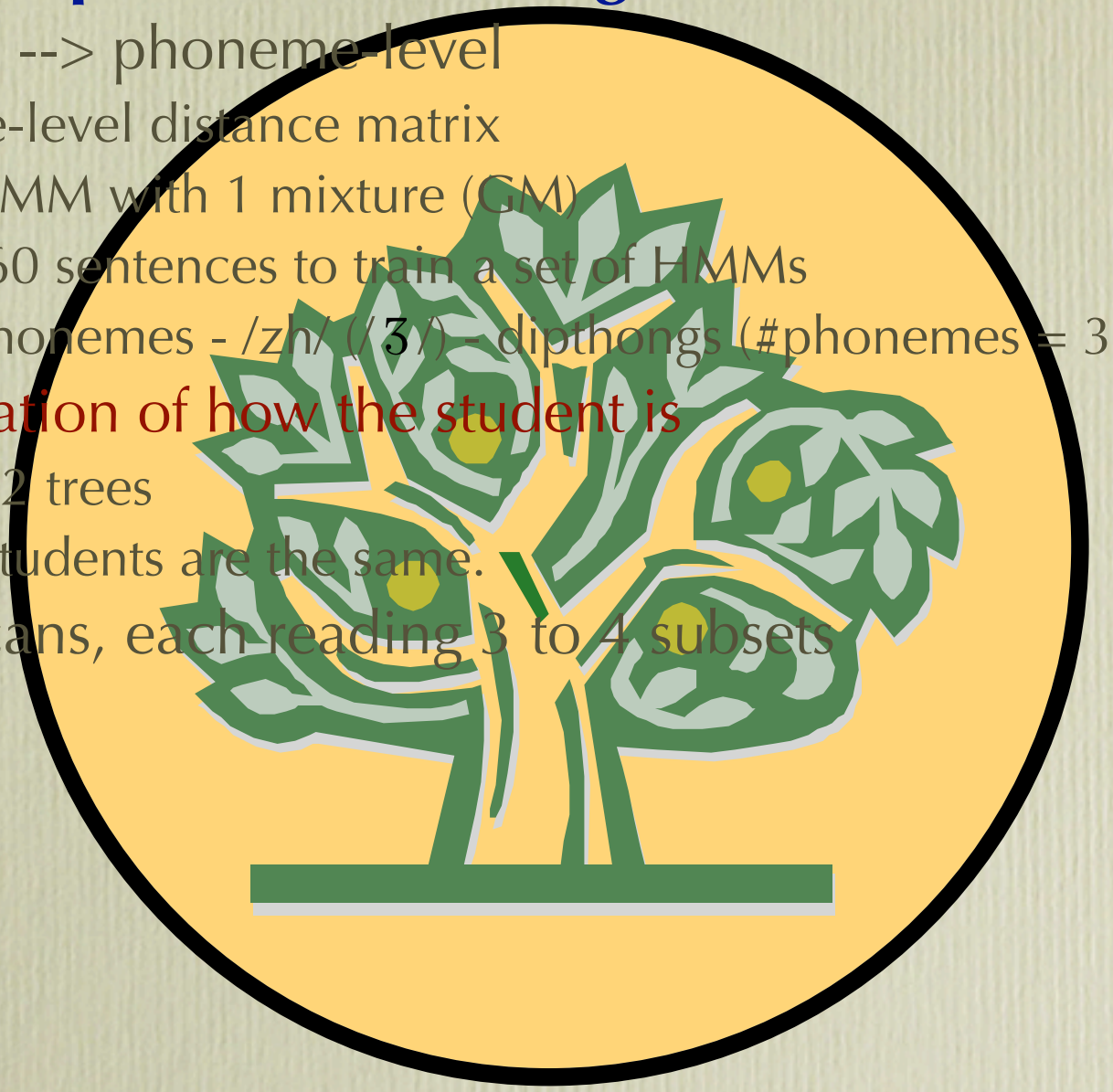
Phonetic interpretation of the tree structure



Corpus analysis of JE production (SD, #1)

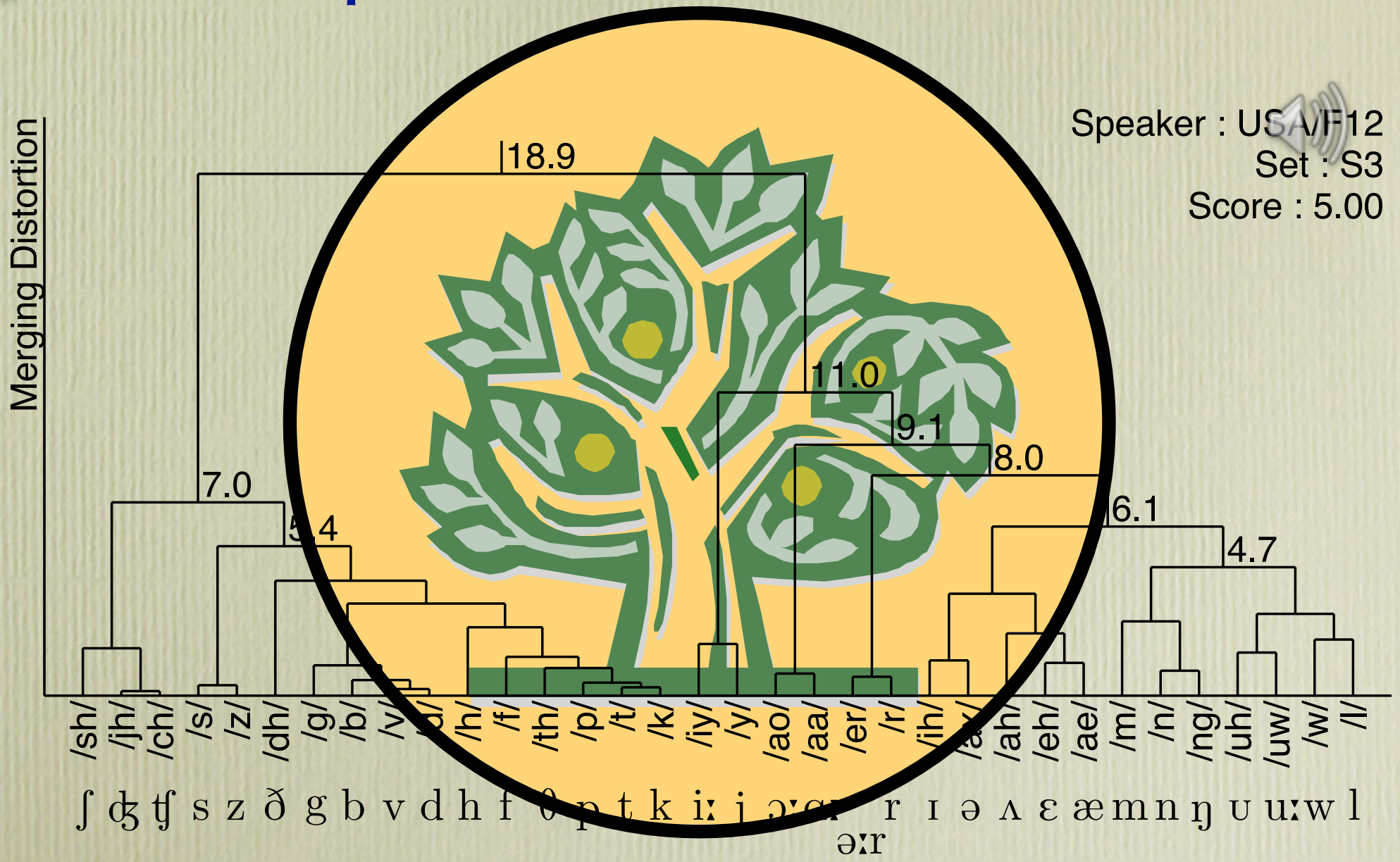
Speaker Dependent tree diagrams

- State-level --> phoneme-level
  - Phoneme-level distance matrix
  - 1-state HMM with 1 mixture (GM)
  - approx. 60 sentences to train a set of HMMs
  - All the phonemes - /zh/ (/ʒ/) - diphthongs (#phonemes = 34)
- Representation of how the student is
  - 100 + 102 trees
  - No two students are the same.
- 20 Americans, each reading 3 to 4 subsets



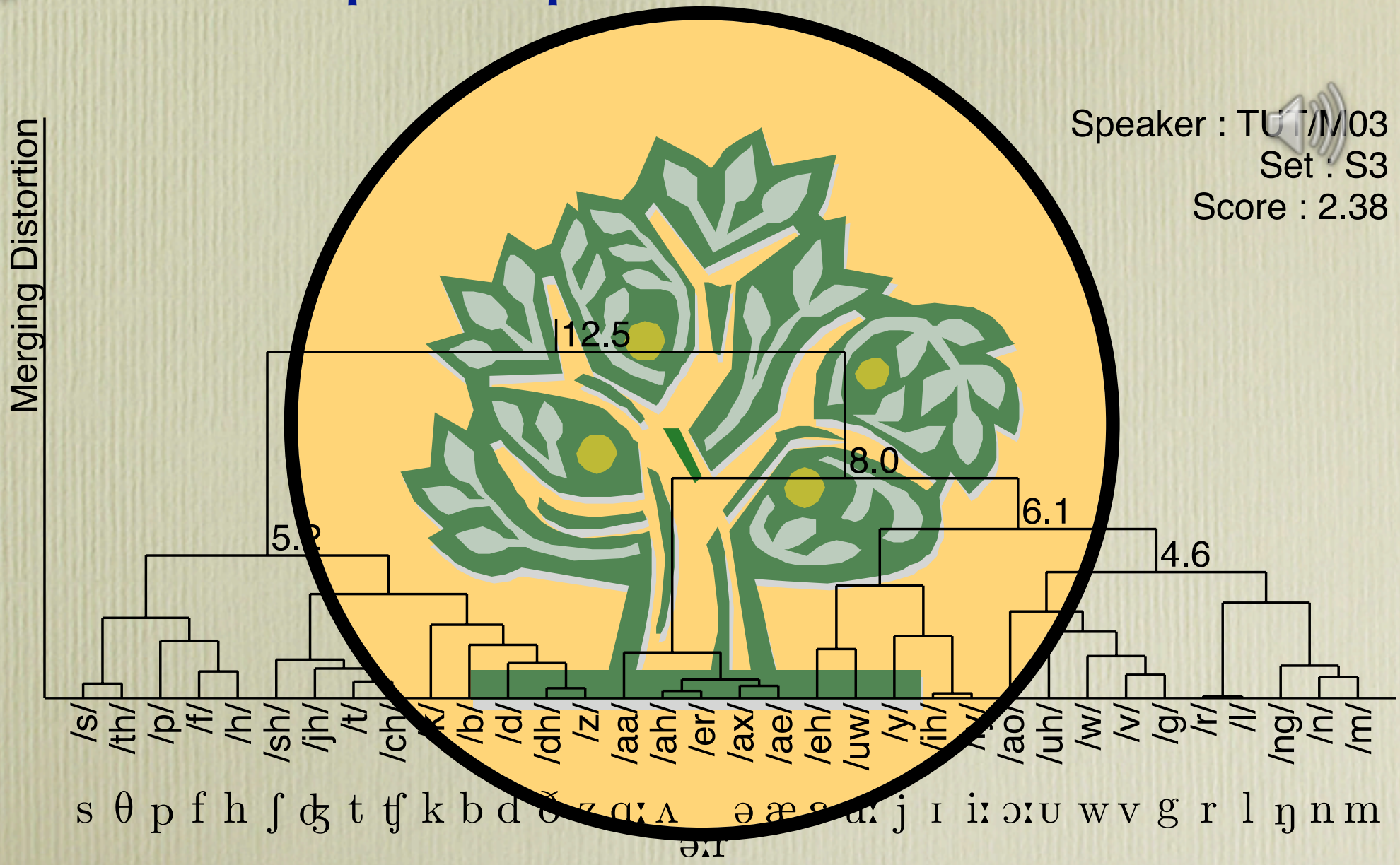
Corpus analysis of JE production (SD, #2)

A PTA example of an American teacher (subset = S3)



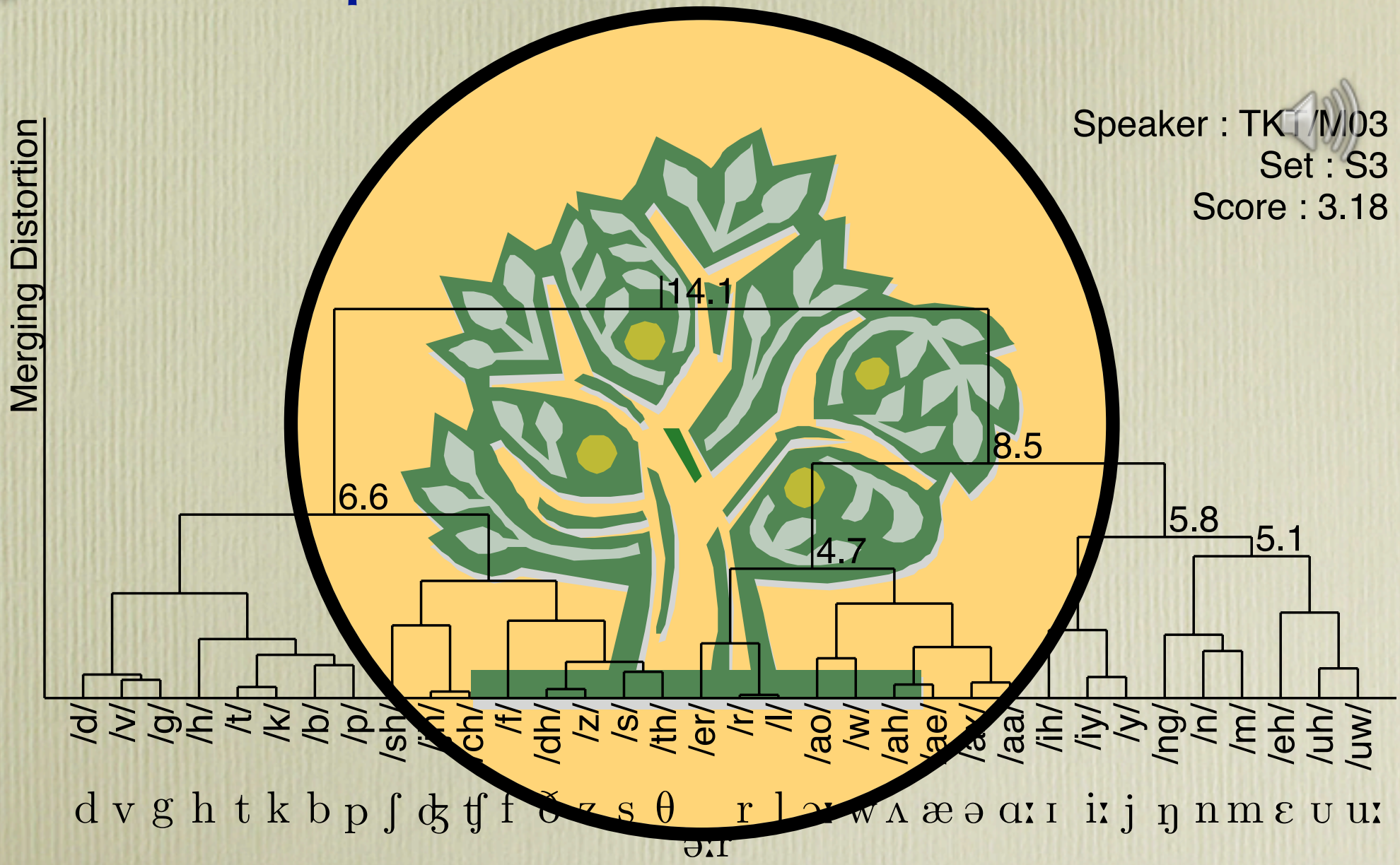
Corpus analysis of JE production (SD, #3)

A PTA example of a poor student (subset = S3)



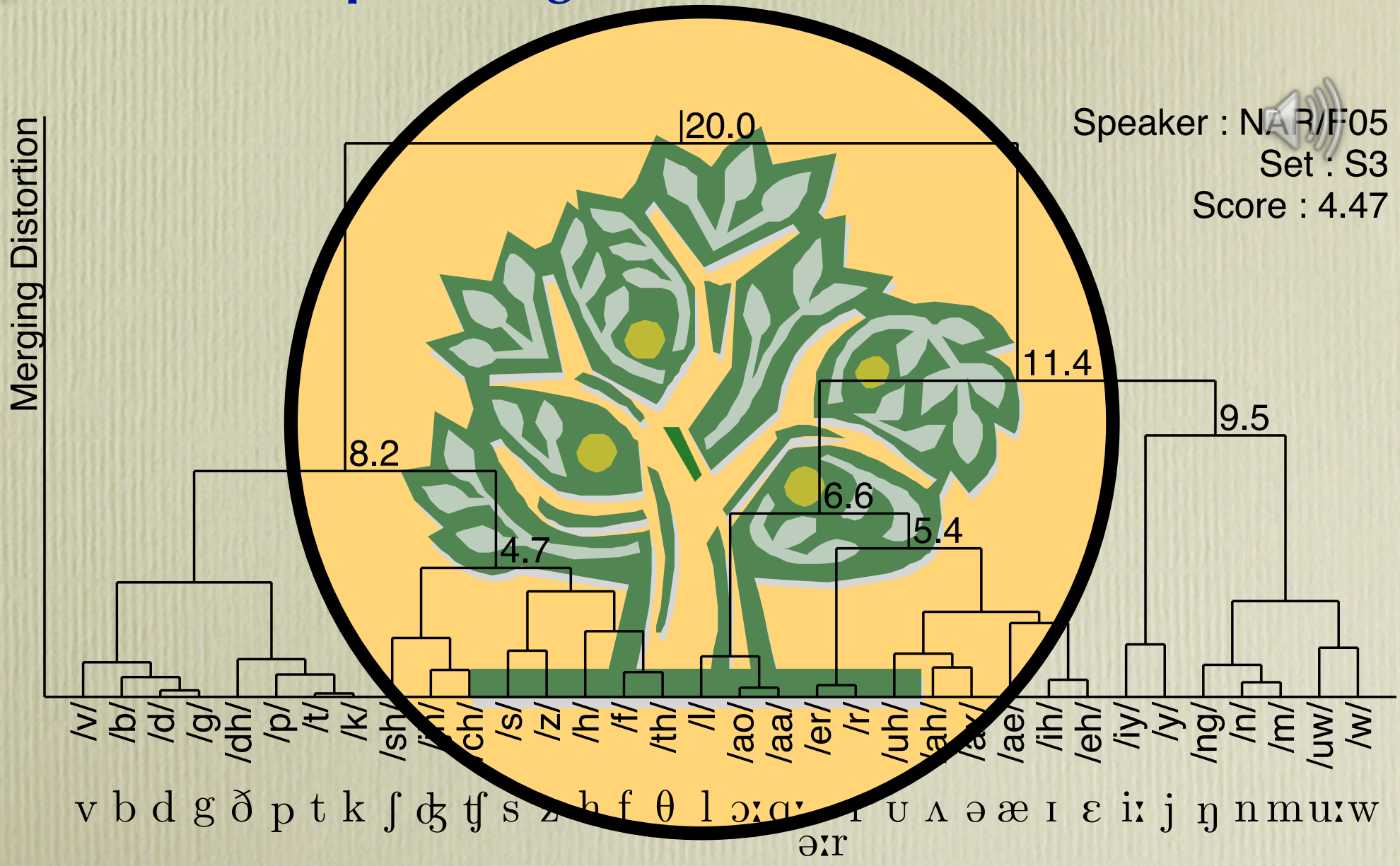
Corpus analysis of JE production (SD, #4)

A PTA example of an intermediate student (subset = S3)



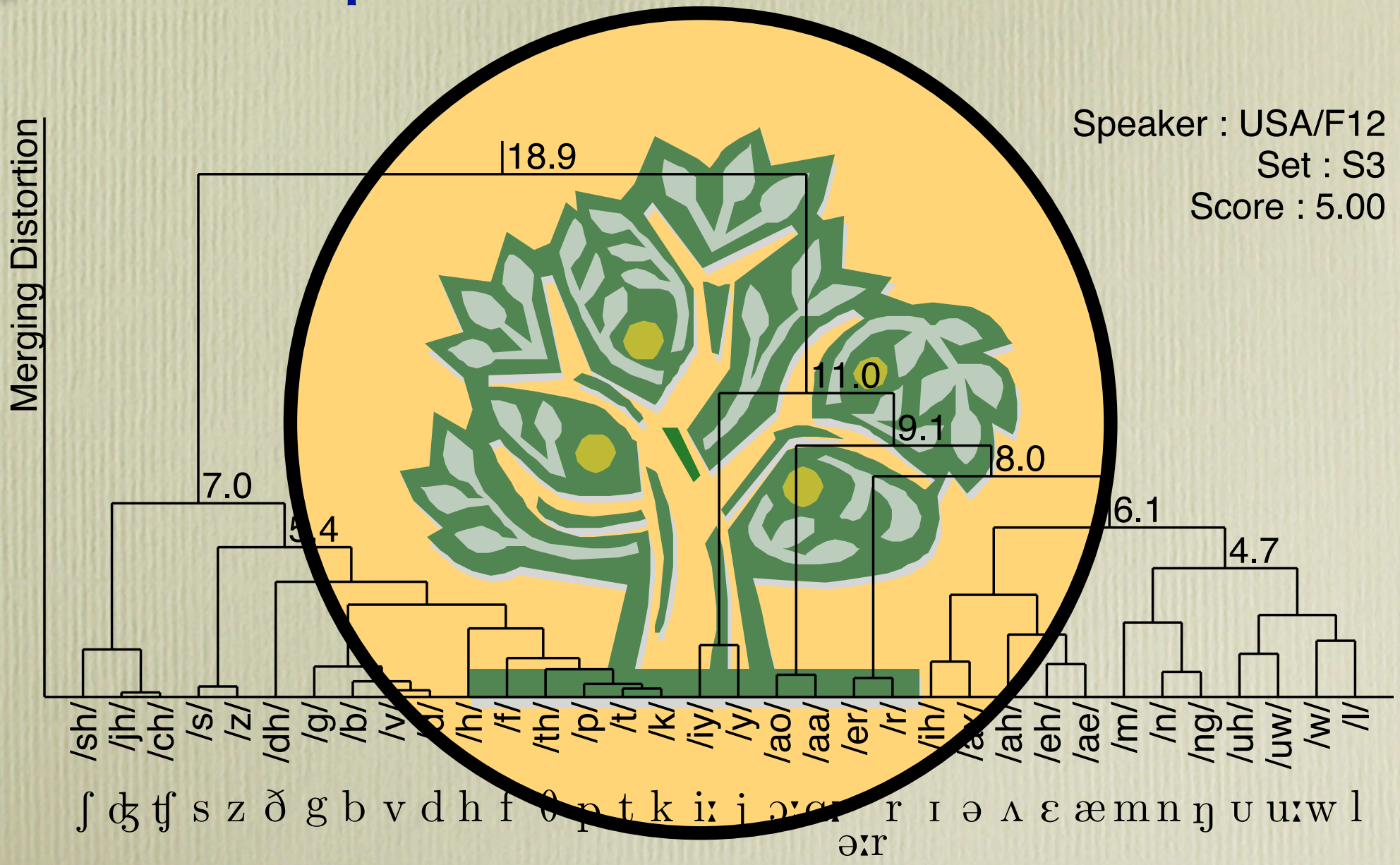
Corpus analysis of JE production (SD, #5)

A PTA example of a good student (subset = S3)



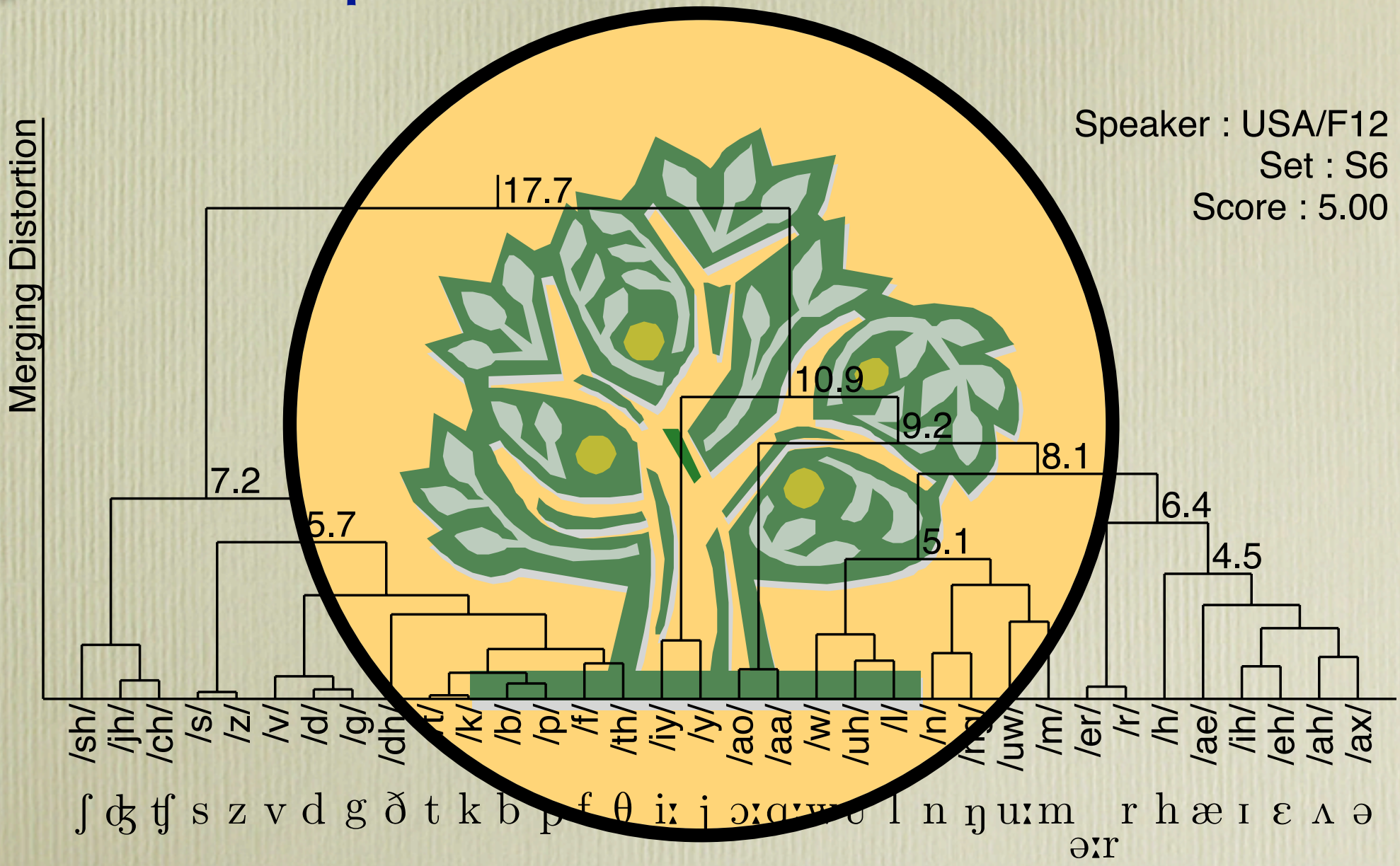
Corpus analysis of JE production (SD, #6)

A PTA example of an American teacher (subset = S3)



Corpus analysis of JE production (SD, #7)

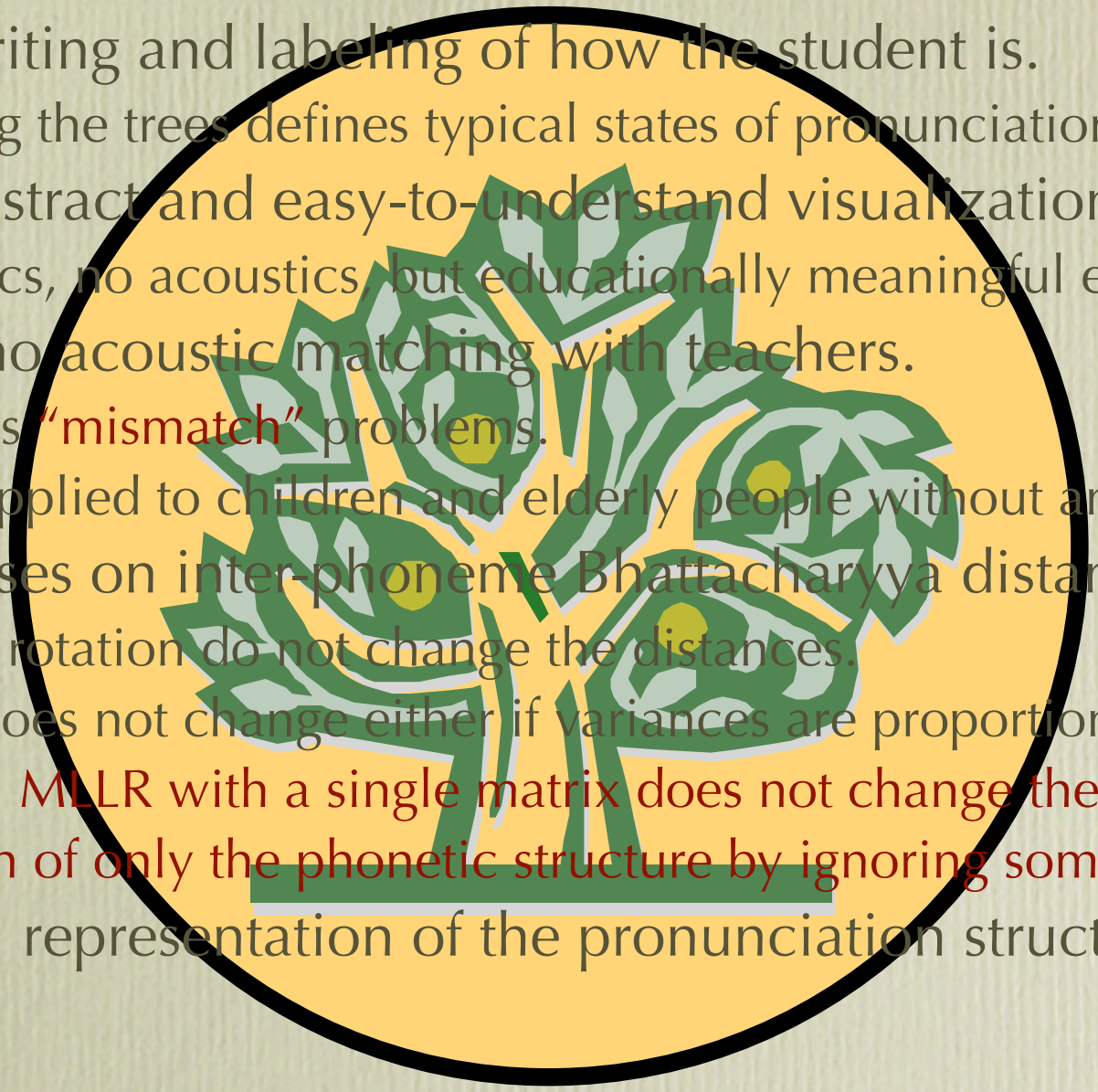
A PTA example of an American teacher (subset = S6)



Corpus analysis of JE production (SD, #8)

Characteristics of PTA

- Can do writing and labeling of how the student is.
  - Clustering the trees defines typical states of pronunciation learning.
- Can do abstract and easy-to-understand visualization.
  - No physics, no acoustics, but educationally meaningful enough
- Requires no acoustic matching with teachers.
  - Never has "mismatch" problems.
  - Can be applied to children and elderly people without any difficulty.
- Only focuses on inter-phoneme Bhattacharyya distances.
  - Shift and rotation do not change the distances.
  - Scaling does not change either if variances are proportionally modified.
  - A part of MLLR with a single matrix does not change the distances.
  - Extraction of only the phonetic structure by ignoring some other factors.
- Perceptual representation of the pronunciation structure.



Corpus analysis of JE production (SD, #9)

Labeling of phonemes

- Phonemic transcriptions generated by forced alignment
- Sequence of phonemes seen in speech of some seconds

Labeling of rhythms

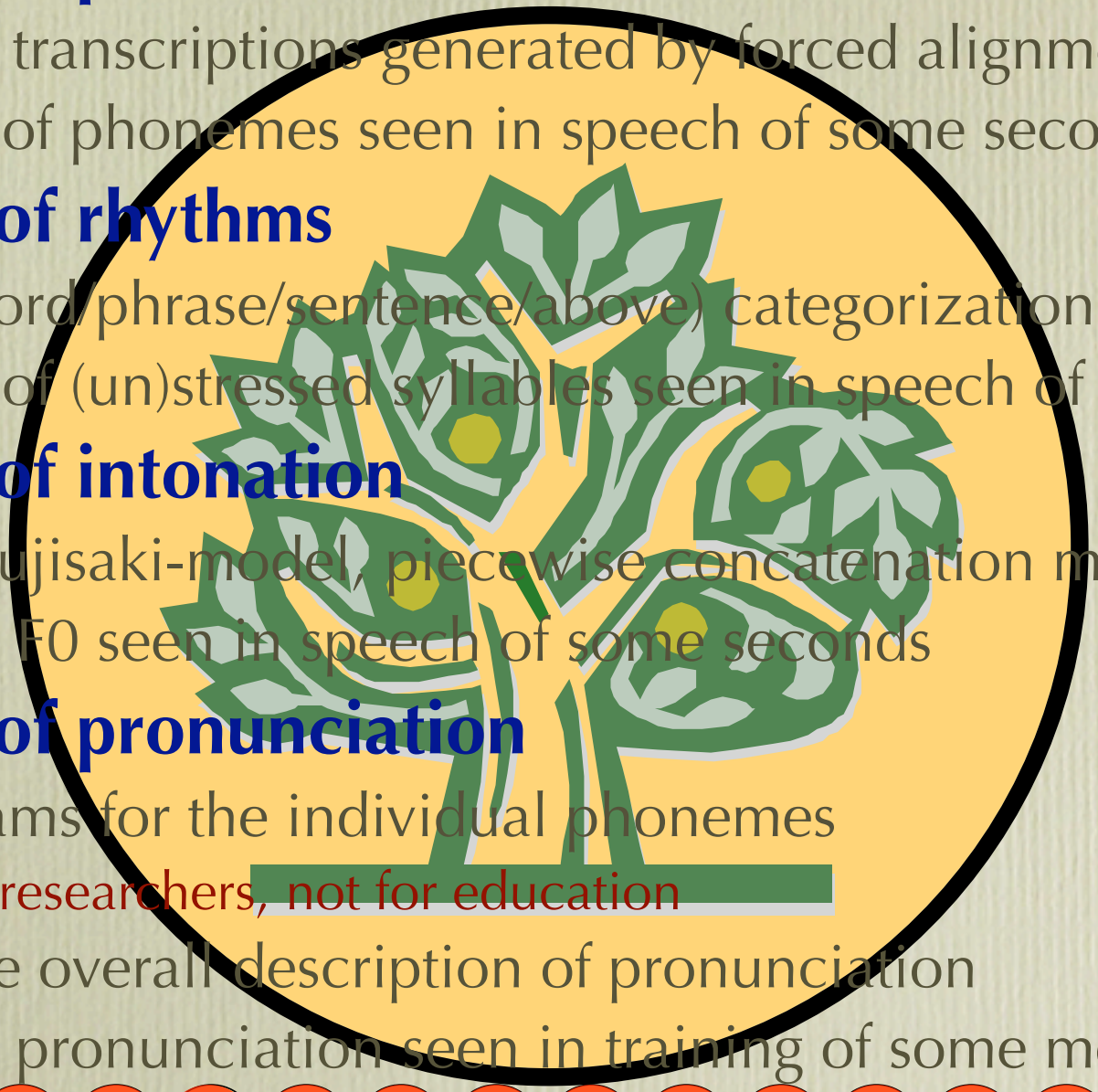
- 4-level (word/phrase/sentence/above) categorization of stress
- Sequence of (un)stressed syllables seen in speech of some seconds

Labeling of intonation

- (x-)ToBI, Fujisaki-model, piecewise concatenation model,,,
- Change of F0 seen in speech of some seconds

Labeling of pronunciation

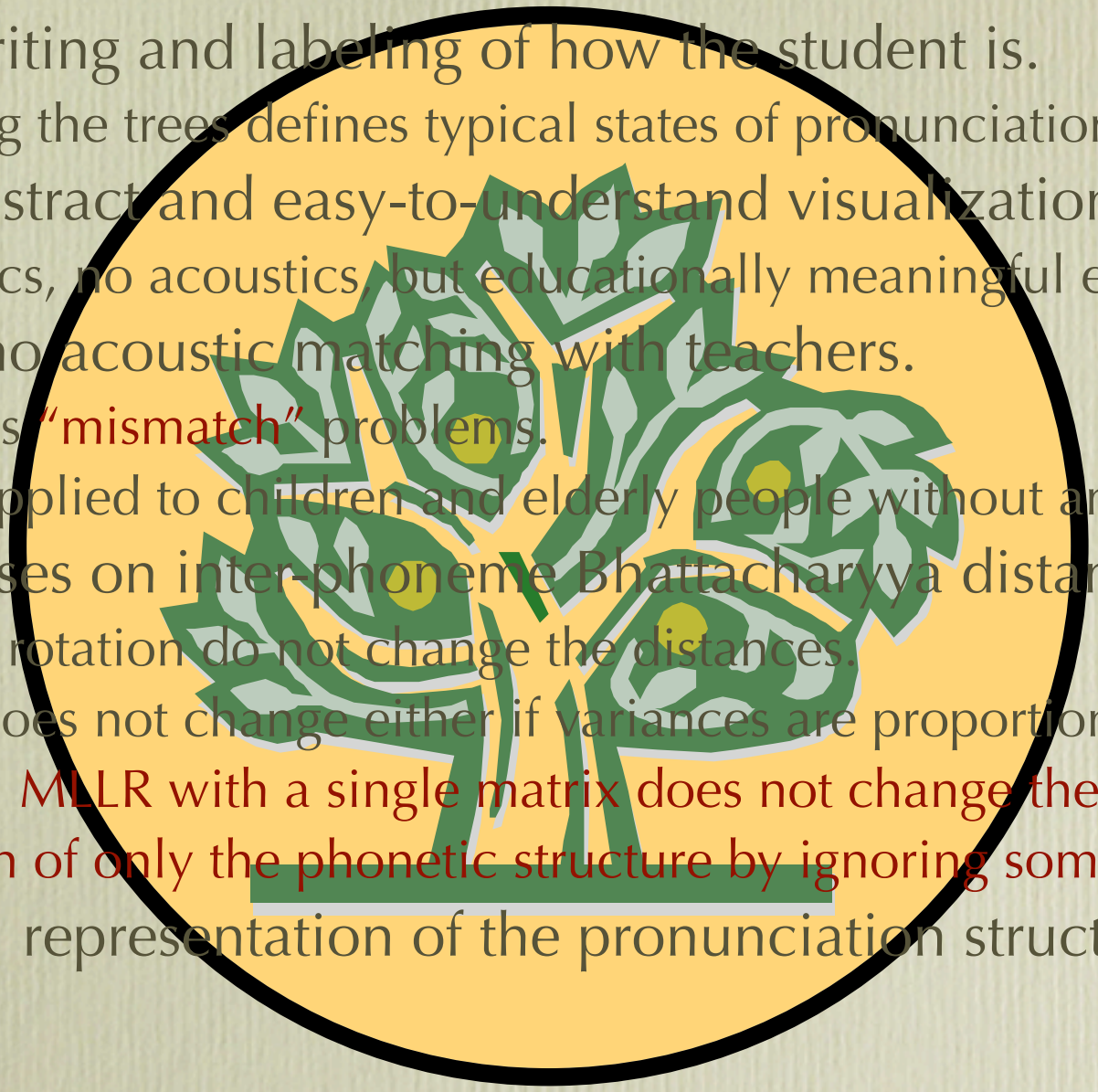
- Spectrograms for the individual phonemes
  - Only for researchers, not for education
- PTA for the overall description of pronunciation
- Change of pronunciation seen in training of some months or years



Corpus analysis of JE production (SD, #8)

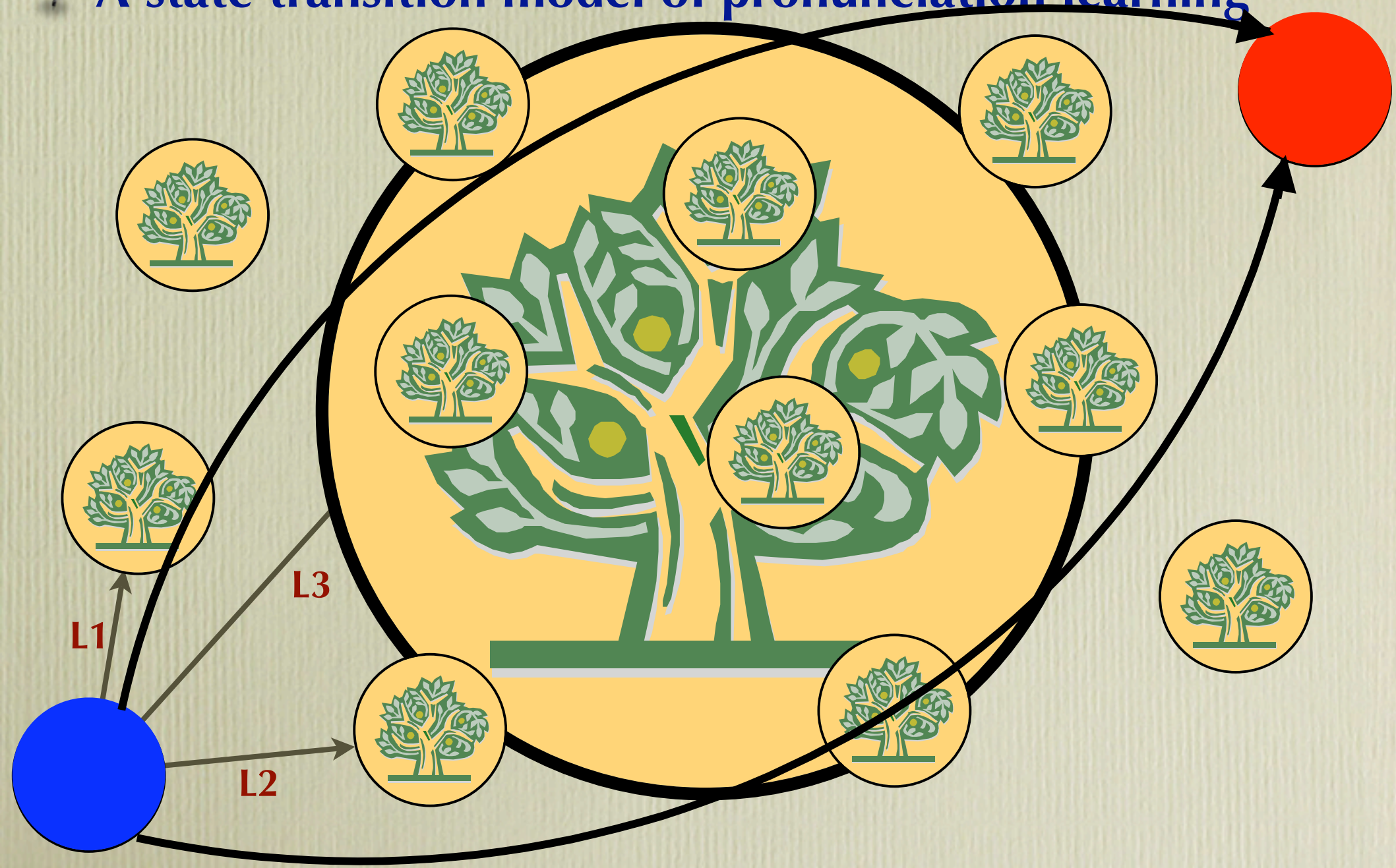
Characteristics of PTA

- Can do writing and labeling of how the student is.
  - Clustering the trees defines typical states of pronunciation learning.
- Can do abstract and easy-to-understand visualization.
  - No physics, no acoustics, but educationally meaningful enough
- Requires no acoustic matching with teachers.
  - Never has "mismatch" problems.
  - Can be applied to children and elderly people without any difficulty.
- Only focuses on inter-phoneme Bhattacharyya distances.
  - Shift and rotation do not change the distances.
  - Scaling does not change either if variances are proportionally modified.
  - A part of MLLR with a single matrix does not change the distances.
  - Extraction of only the phonetic structure by ignoring some other factors.
- Perceptual representation of the pronunciation structure.



Corpus analysis of JE production (SD, #10)

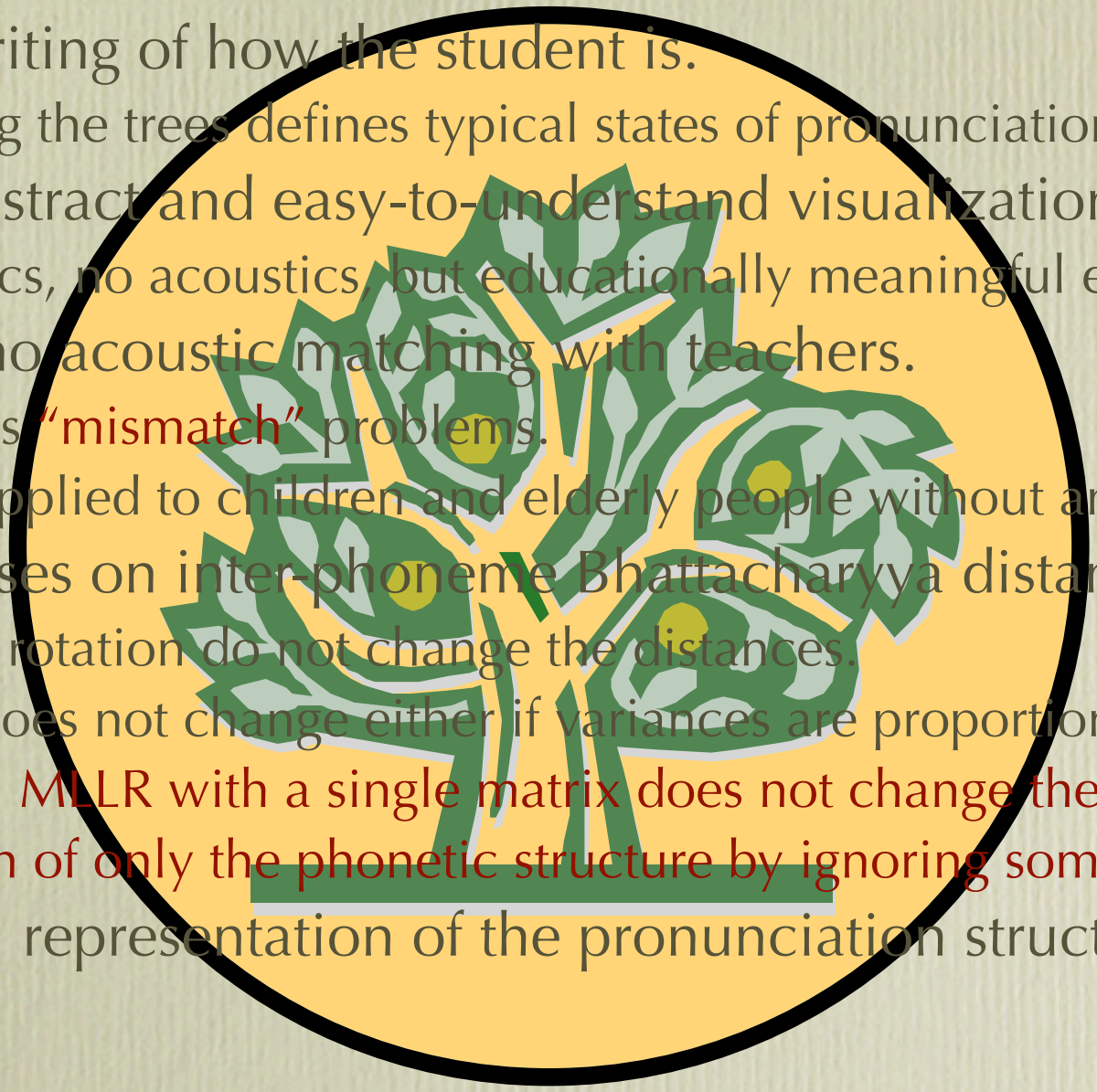
A state transition model of pronunciation learning



Corpus analysis of JE production (SD, #8)

Characteristics of PTA

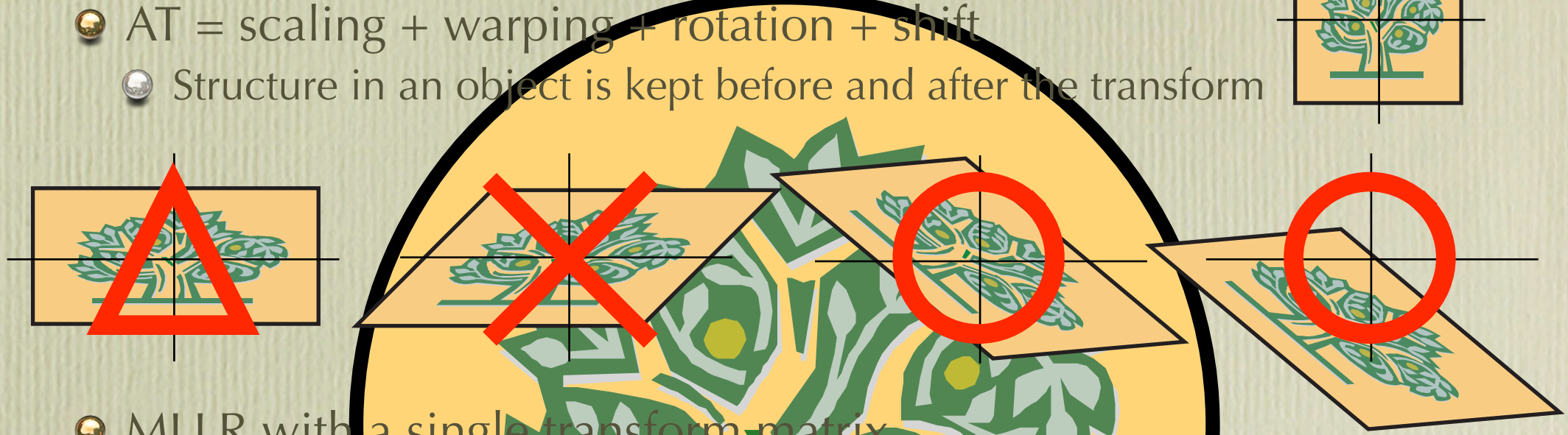
- Can do writing of how the student is.
  - Clustering the trees defines typical states of pronunciation learning.
- Can do abstract and easy-to-understand visualization.
  - No physics, no acoustics, but educationally meaningful enough
- Requires no acoustic matching with teachers.
  - Never has "mismatch" problems.
  - Can be applied to children and elderly people without any difficulty.
- Only focuses on inter-phoneme Bhattacharyya distances.
  - Shift and rotation do not change the distances.
  - Scaling does not change either if variances are proportionally modified.
  - A part of MLLR with a single matrix does not change the distances.
  - Extraction of only the phonetic structure by ignoring some other factors.
- Perceptual representation of the pronunciation structure.



Corpus analysis of JE production (SD, #11)

MLLR = Affine Transform of cepstrums

- AT = scaling + warping + rotation + shift
  - Structure in an object is kept before and after the transform

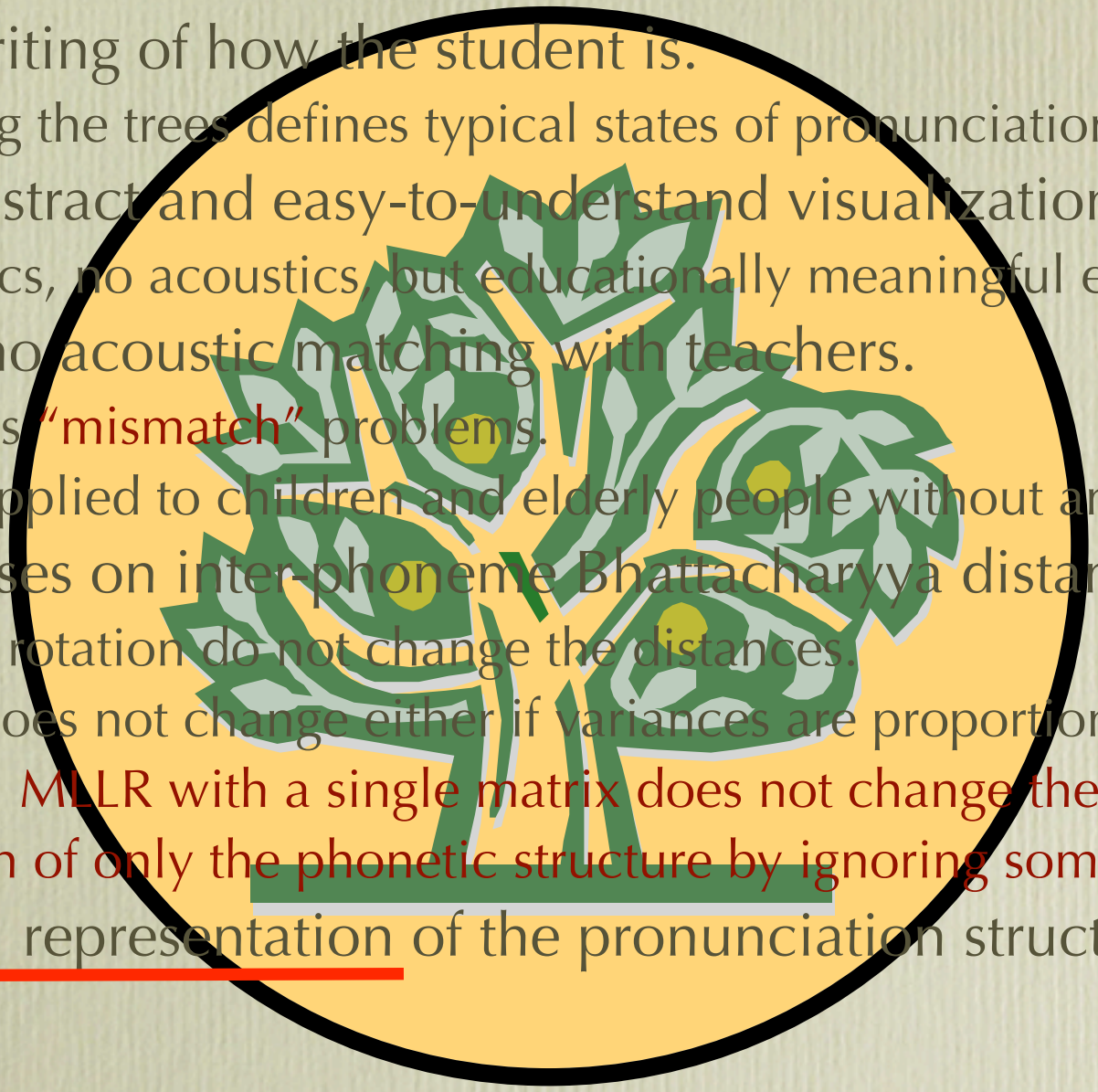


- MLLR with a single transform matrix
- MLLR adaptation with N (>>1) matrices in speech recognition
  - Different parts of a triphone set have different speaker individuality.
- Speaker individuality can be modeled by a single GMM.
- MLLR adaptation in HMM-based speech synthesis
  - HMMs of a speaker are converted into those of another with 5 sentences.
  - N ~ 1
- Good cancellation of speaker individuality

Corpus analysis of JE production (SD, #8)

Characteristics of PTA

- Can do writing of how the student is.
  - Clustering the trees defines typical states of pronunciation learning.
- Can do abstract and easy-to-understand visualization.
  - No physics, no acoustics, but educationally meaningful enough
- Requires no acoustic matching with teachers.
  - Never has "mismatch" problems.
  - Can be applied to children and elderly people without any difficulty.
- Only focuses on inter-phoneme Bhattacharyya distances.
  - Shift and rotation do not change the distances.
  - Scaling does not change either if variances are proportionally modified.
  - A part of MLLR with a single matrix does not change the distances.
  - Extraction of only the phonetic structure by ignoring some other factors.
- Perceptual representation of the pronunciation structure.



Is PTA new ? (#1)

NO !

- "Structuralism" by R. Jakobson, M. Halle, G. Fant, and etc
  - Concise and accurate description of sounds in a language
  - Phoneme clustering based upon distinctive features

Is PTA new ? (#2)

"The Sound Pattern of Russian" by M. Halle

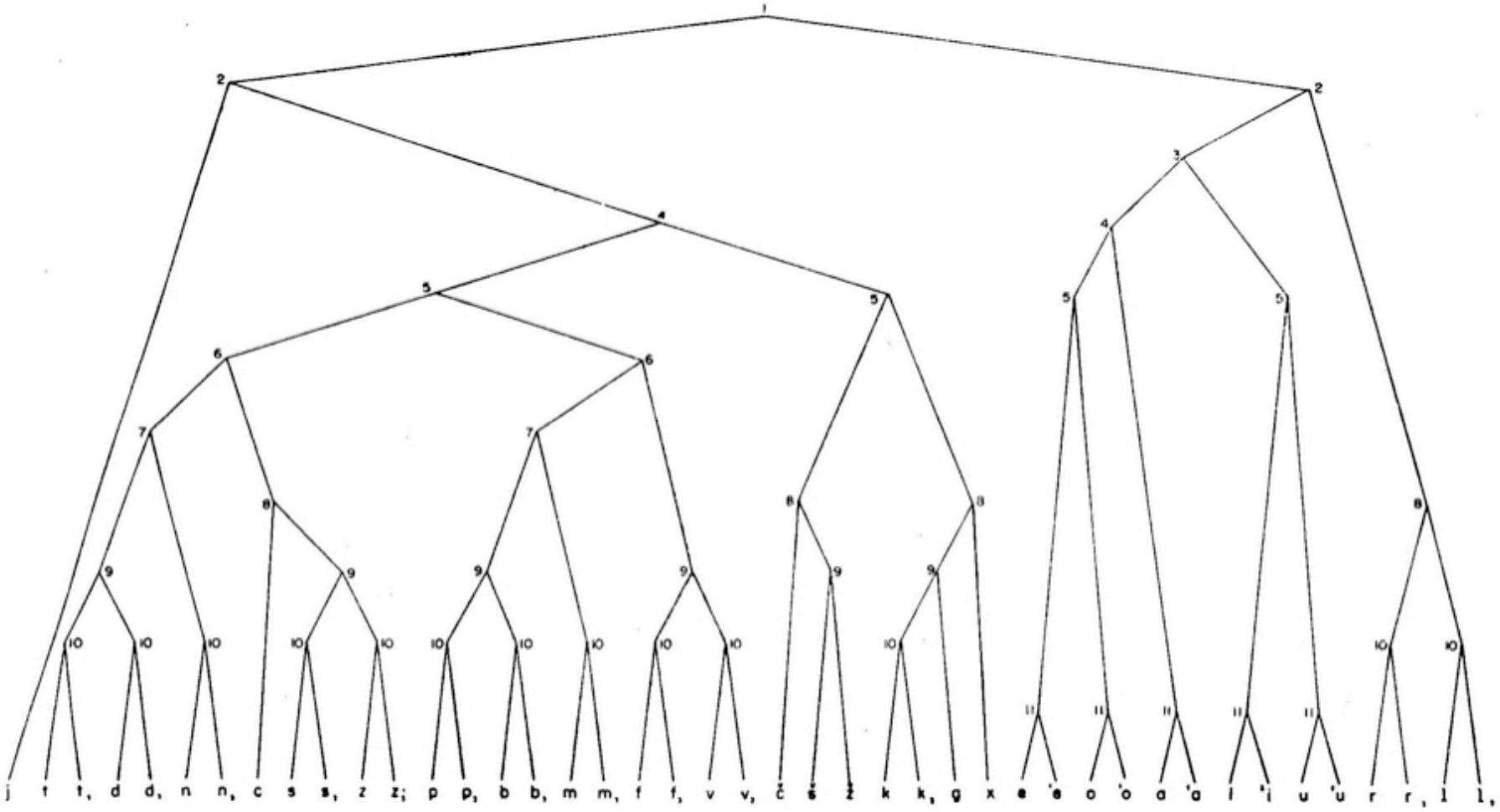


Fig. I-1. Branching diagram representing the morphonemes of Russian. The numbers with which each node is labelled refer to the different features, as follows: 1. vocalic vs. nonvocalic; 2. consonantal vs. nonconsonantal; 3. diffuse vs. nondiffuse; 4. compact vs. noncompact; 5. low tonality vs. high tonality; 6. strident vs. mellow; 7. nasal vs. nonnasal; 8. continuant vs. interrupted; 9. voiced vs. voiceless; 10. sharped vs. plain; 11. accented vs. unaccented. Left branches represent minus values, and right branches, plus values for the particular feature.

Is PTA new ? (#1)

NO !

- "Structuralism" by R. Jakobson, M. Halle, G. Fant, and etc
  - Concise and accurate description of sounds in a language
  - Phoneme clustering based upon distinctive features

YES !

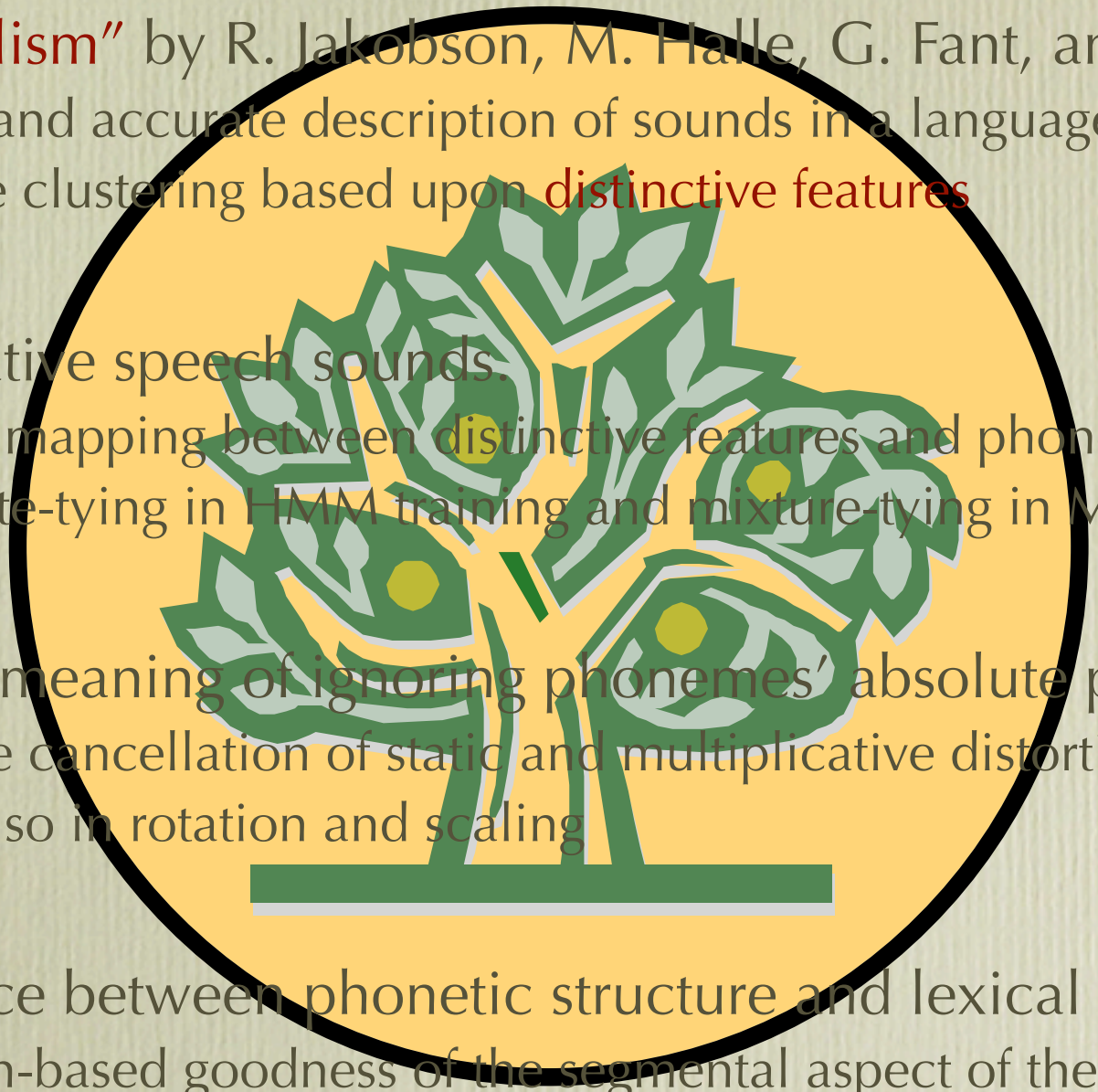
- It's non-native speech sounds.
  - No fixed mapping between distinctive features and phonemes
  - Some state-tying in HMM training and mixture-tying in MLLR adaptation

YES !!

- Technical meaning of ignoring phonemes' absolute positions in AS
  - Complete cancellation of static and multiplicative distortions
  - Robust also in rotation and scaling

YES !!!

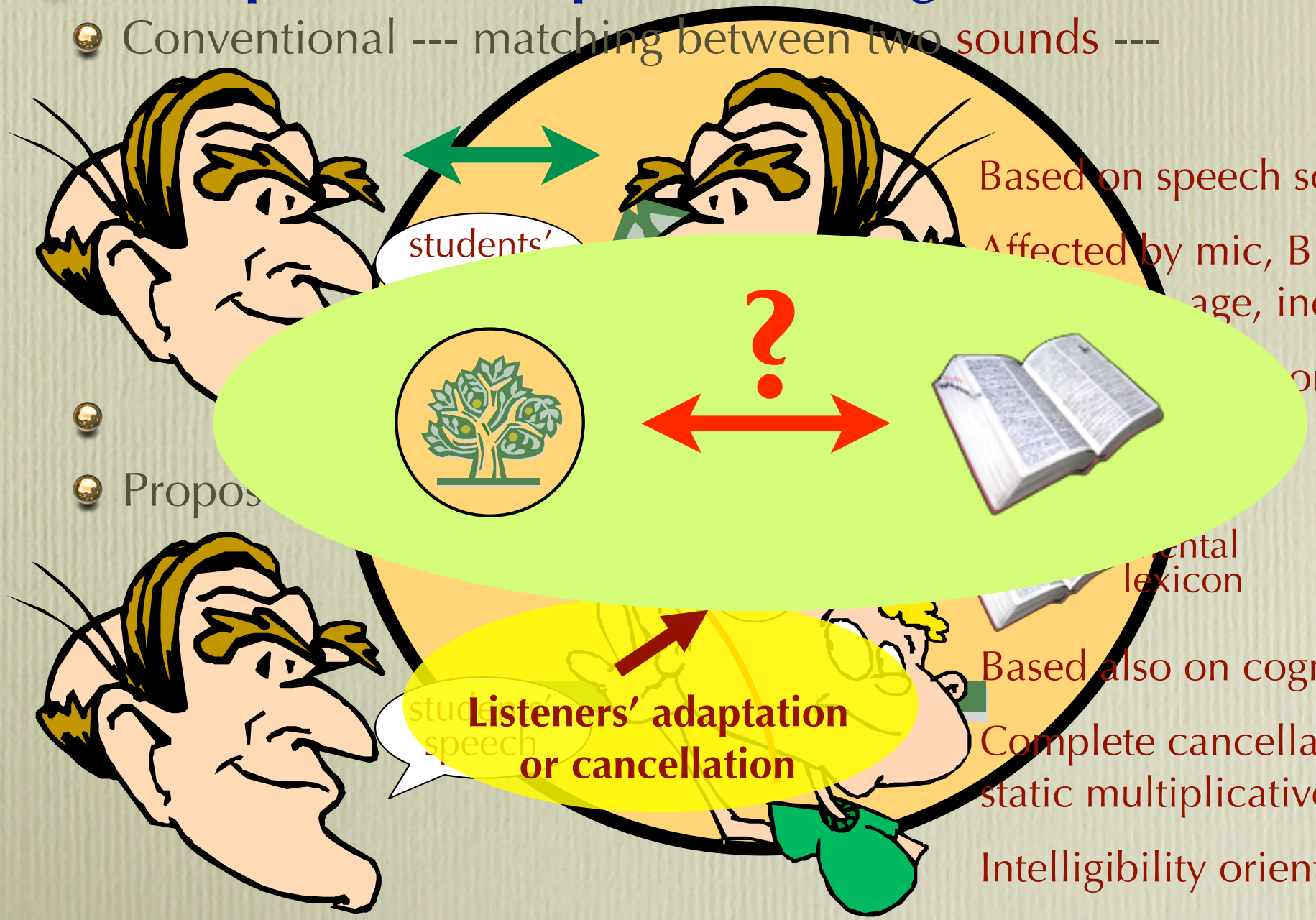
- Accordance between phonetic structure and lexical structure
  - Cognition-based goodness of the segmental aspect of the pronunciation
  - Not native sounding but intelligibility-based scoring



Corpus analysis of JE perception (SI, #1)

From phonetics to phonetics+cognition

- Conventional --- matching between two sounds ---



Based on speech sci. and eng.

Affected by mic, BN, size, age, individuality

oriented

- Propos

ental lexicon

Based also on cognition

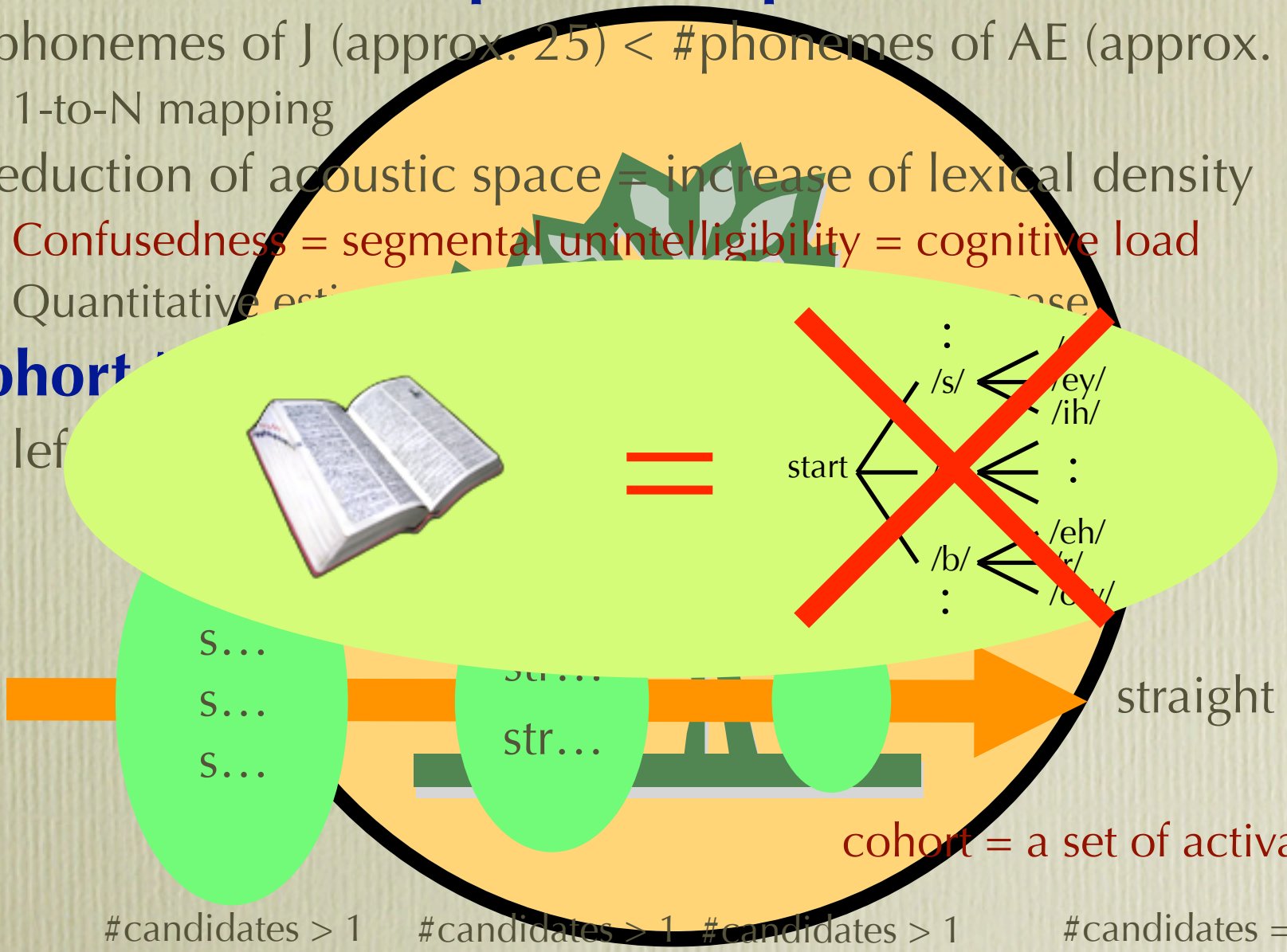
Complete cancellation of static multiplicative noise

Intelligibility oriented

Corpus analysis of JE perception (SI, #2)

Reduced acoustic space of JE phonemes

- #phonemes of J (approx. 25) < #phonemes of AE (approx. 40)
  - 1-to-N mapping
- Reduction of acoustic space = increase of lexical density
  - Confusedness = segmental unintelligibility = cognitive load
  - Quantitative esti… …ase
- **Cohort …**
  - A lef…



- Estimation of the cohort size with the initial portion of input speech

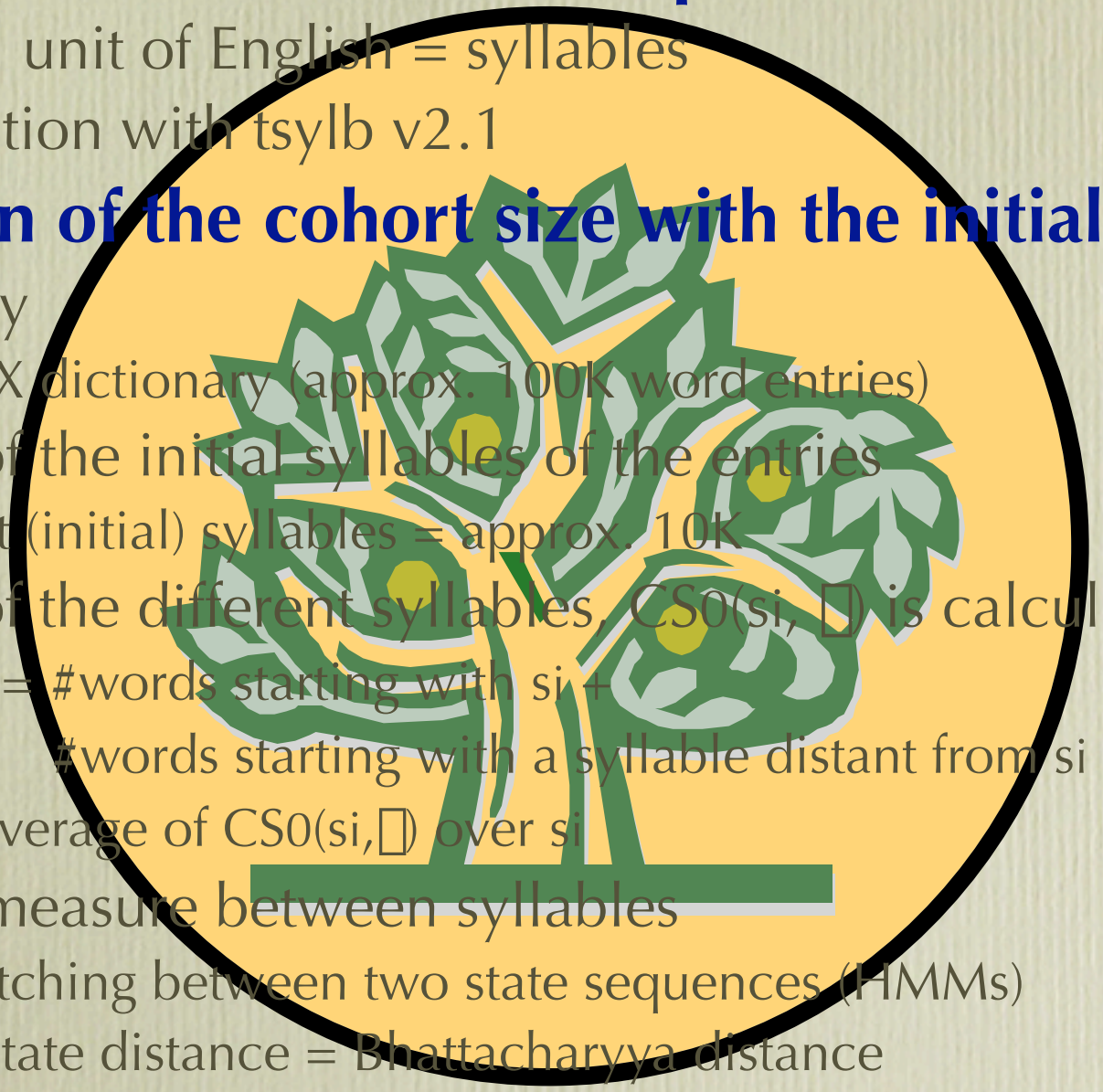
Corpus analysis of JE perception (SI, #3)

Acoustic unit of cohort development

- Perceptual unit of English = syllables
- Syllabification with tsylb v2.1

Estimation of the cohort size with the initial syl. input

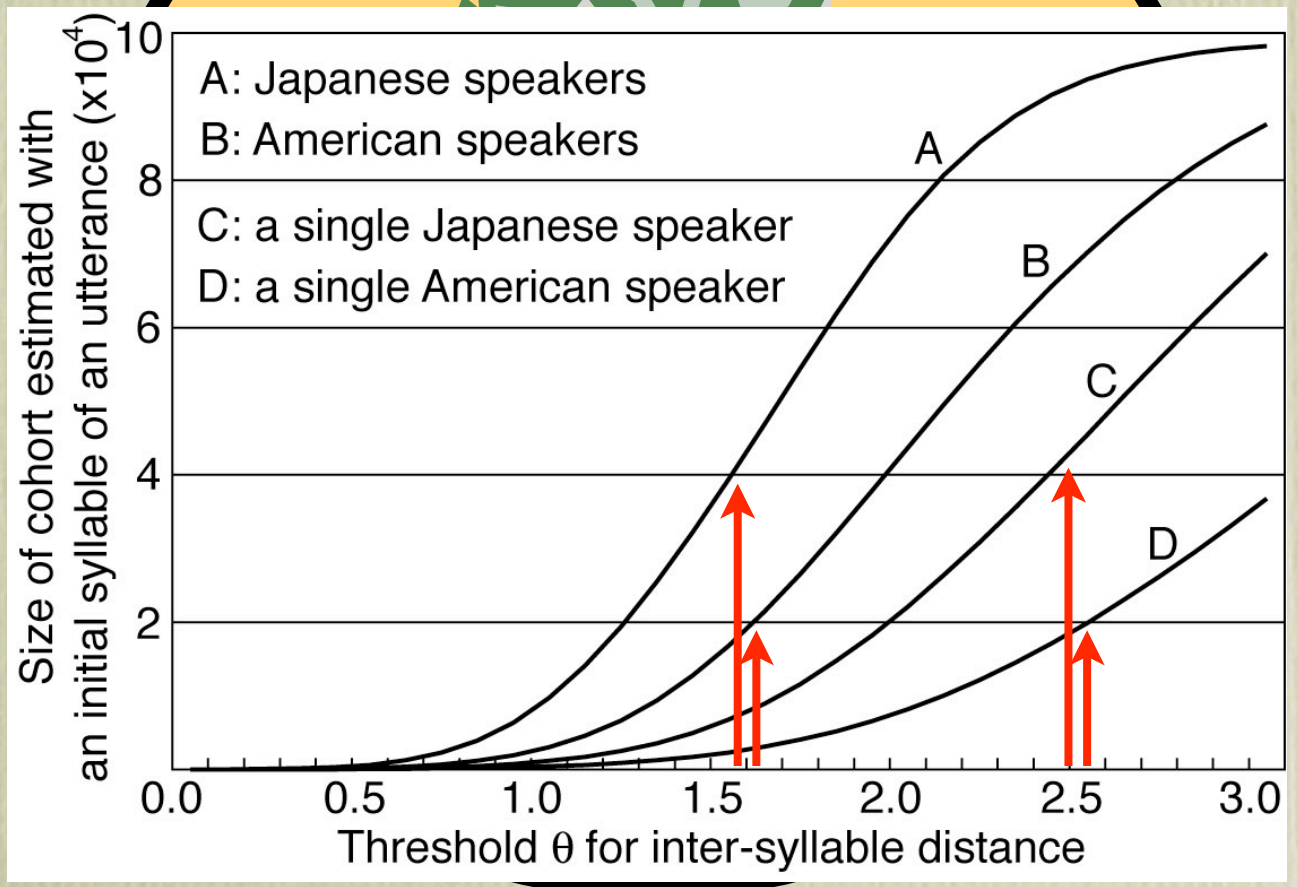
- Vocabulary
  - PRONLEX dictionary (approx. 100K word entries)
- Varieties of the initial syllables of the entries
  - #different (initial) syllables = approx. 10K
- For each of the different syllables, CS0(si, □) is calculated.
  - CS0(si,□) = #words starting with si +
    #words starting with a syllable distant from si by less than □
  - CS(□) = average of CS0(si,□) over si
- Distance measure between syllables
  - = DP matching between two state sequences (HMMs)
  - State-to-state distance = Bhattacharyya distance
  - State-level distance matrix provides all the required information.



Corpus analysis of JE perception (SI, #4)

Cohort size as a function of threshold

- SI AE models vs. SI JE models
- SD AE models vs. SD JE models
- No weighting based on N-gram probabilities



Intelligibility

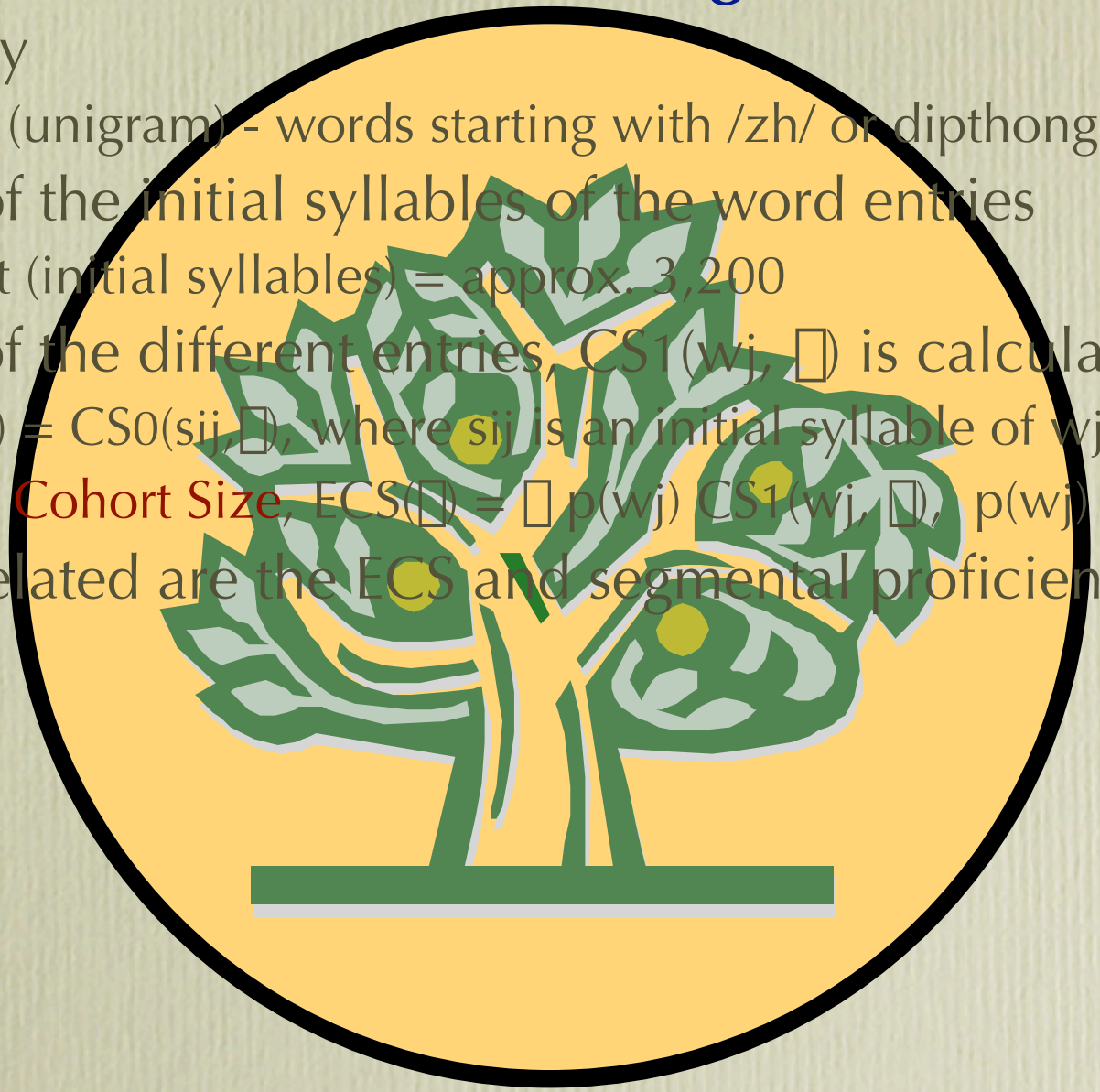
Confusedness

High ← Required resolution of acoustic analysis → Low

Corpus analysis of JE perception (SD, #1)

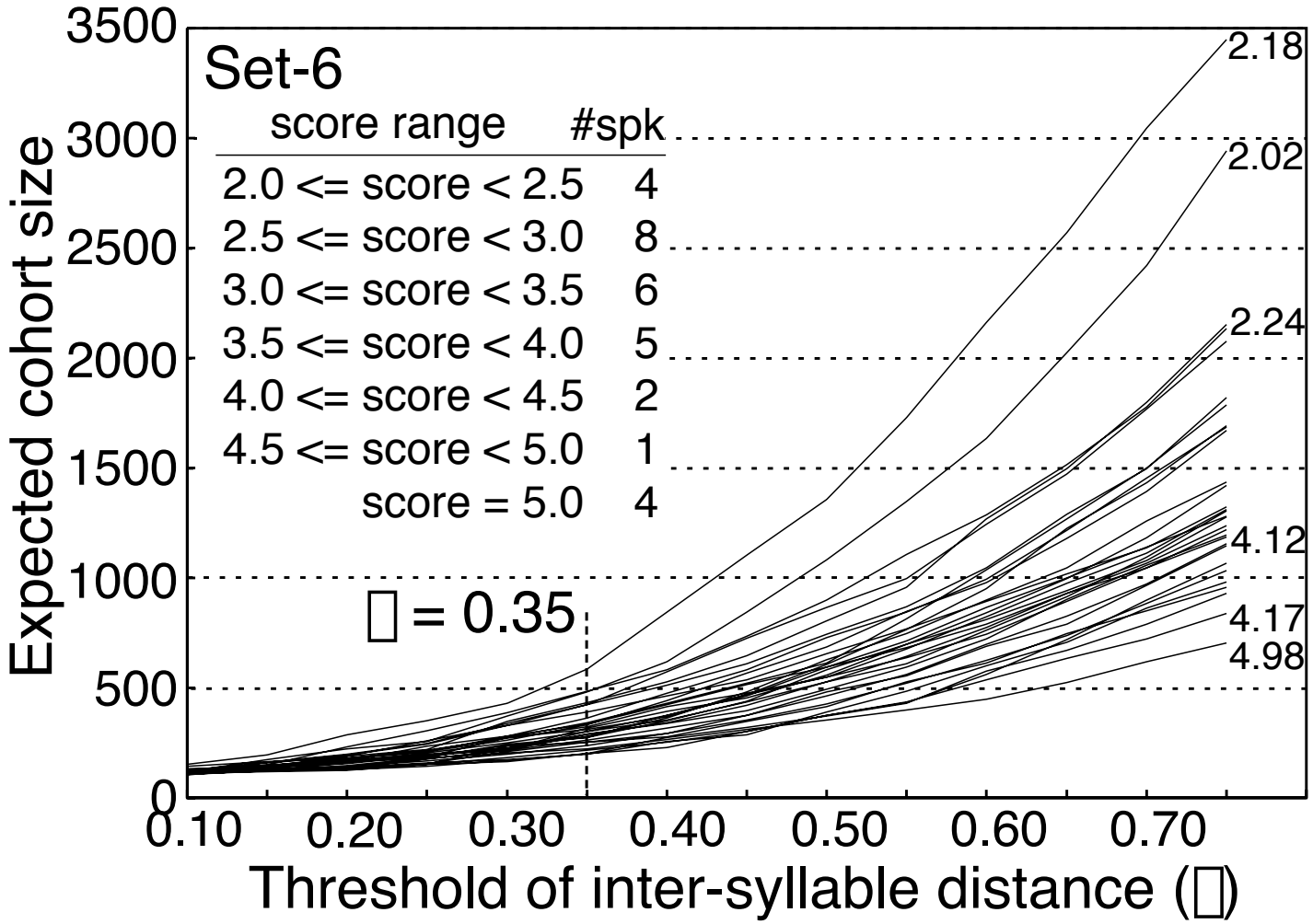
Cohort size estimation with unigram

- Vocabulary
  - WSJ 20K (unigram) - words starting with /zh/ or dipthongs
- Varieties of the initial syllables of the word entries
  - #different (initial syllables) = approx. 3,200
- For each of the different entries, CS1(wj, □) is calculated.
  - CS1(wj,□) = CS0(sij,□), where sij is an initial syllable of wj
  - Expected Cohort Size, ECS(□) = □ p(wj) CS1(wj, □), p(wj) = 1-gram
- How correlated are the ECS and segmental proficiency labels ?



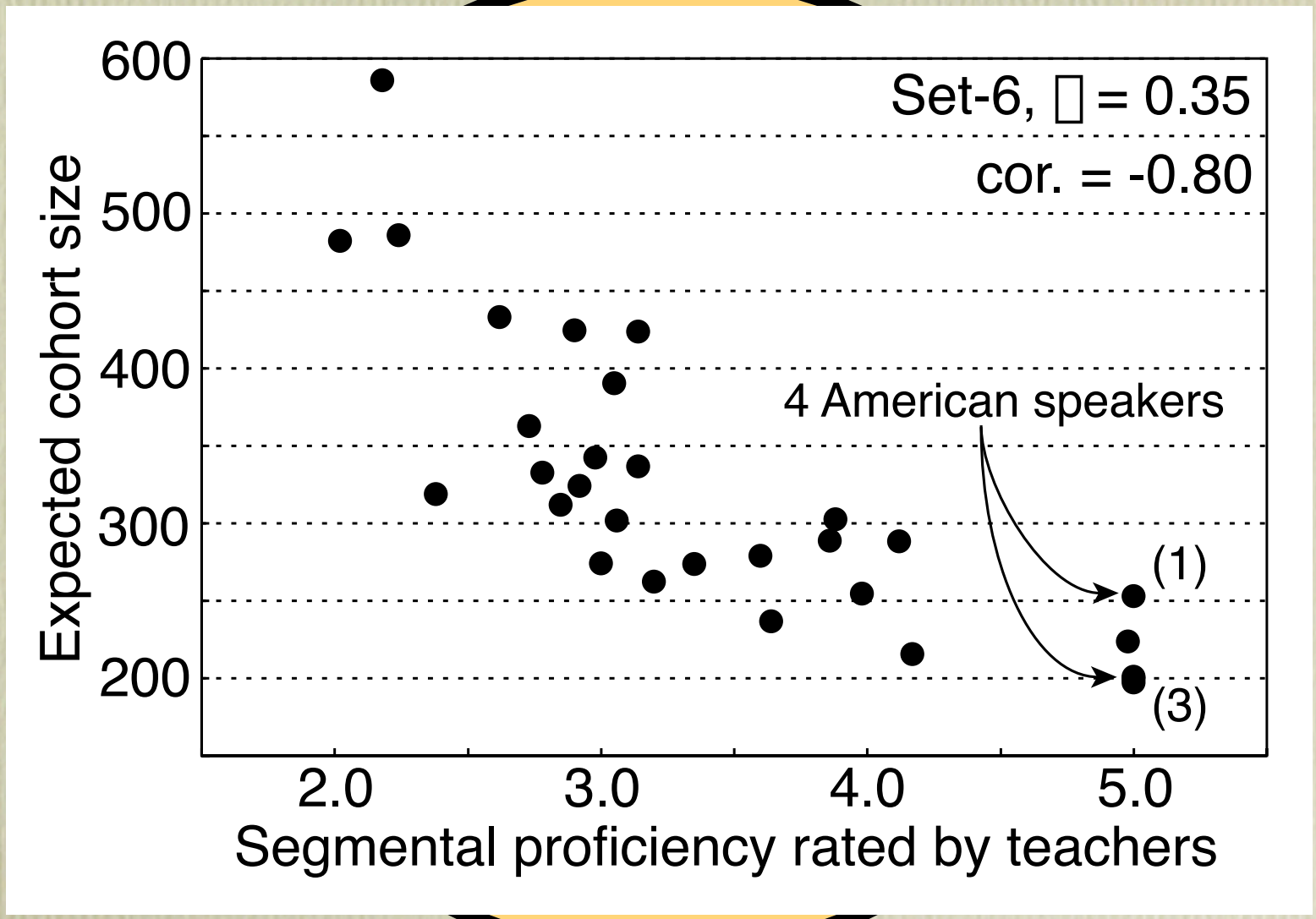
Corpus analysis of JE perception (SD, #2)

Expected cohort size and segmental proficiency



Corpus analysis of JE perception (SD, #3)

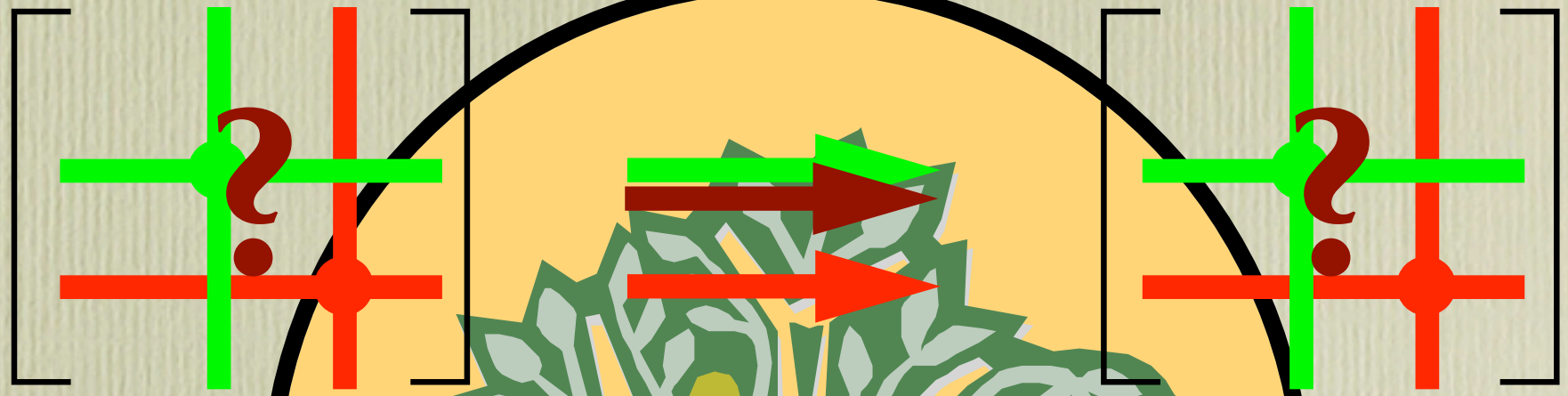
Expected cohort size and segmental proficiency



- Non-acoustic matching can estimate the segmental intelligibility.

Possible applications of PTA (#1)

Estimation of the next target for efficient learning



teacher's matrix　　learner's matrix

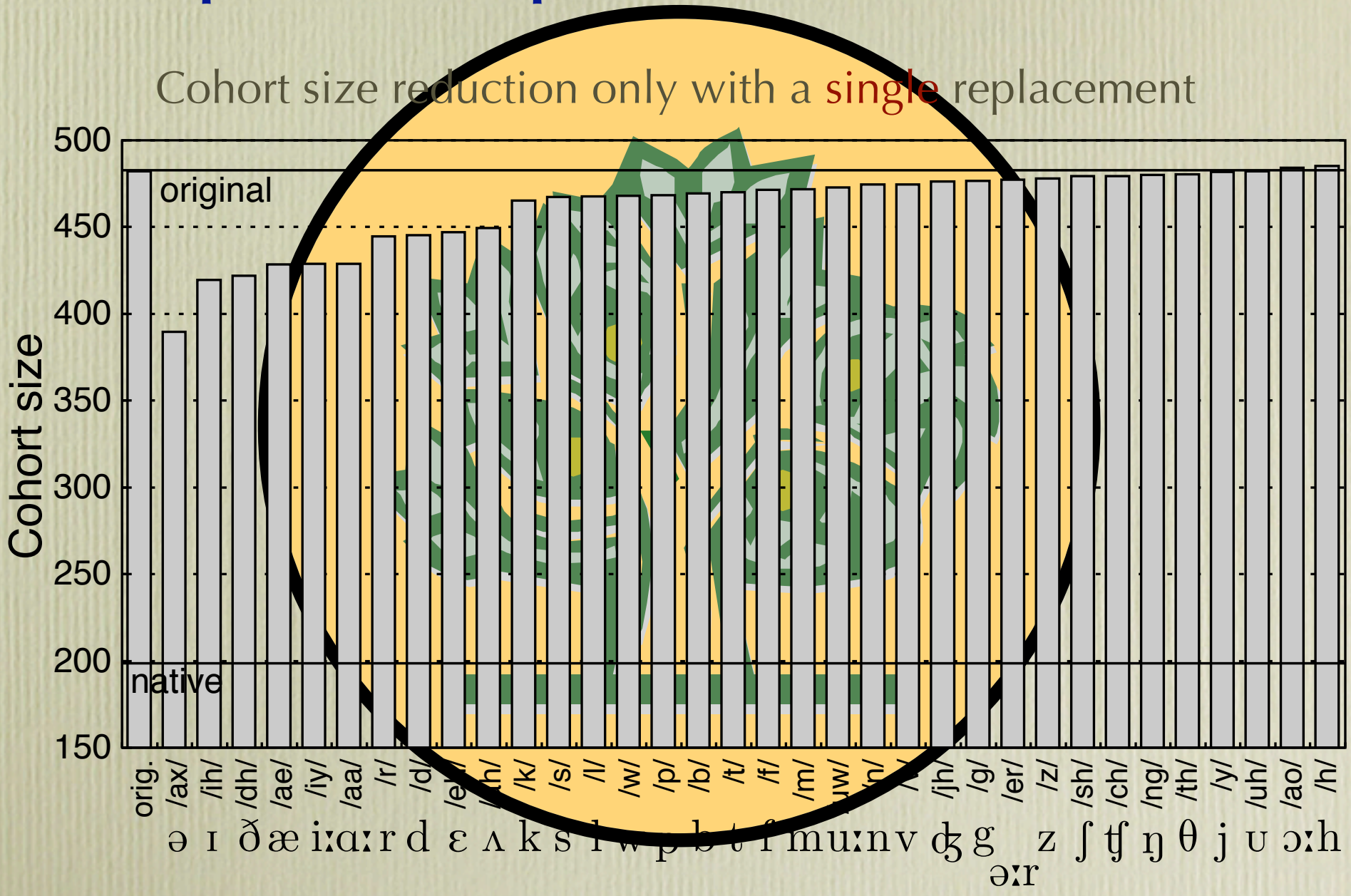
Which replacement realizes the largest reduction of cohort size ?

Speech samples of RYU/F06

- Several sentence speech samples

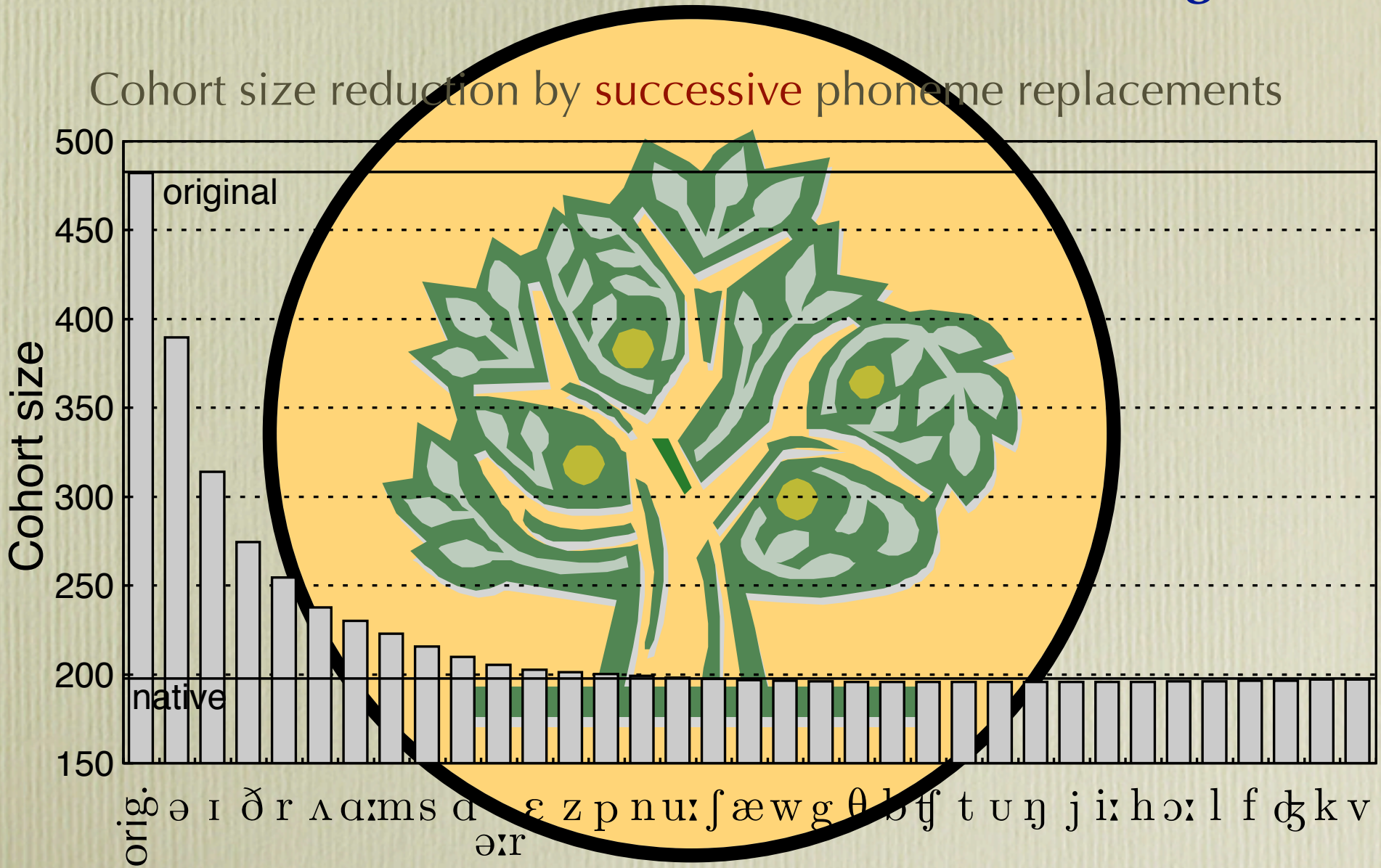
Possible applications of PTA (#2)

Which phoneme's replacement should come first ?



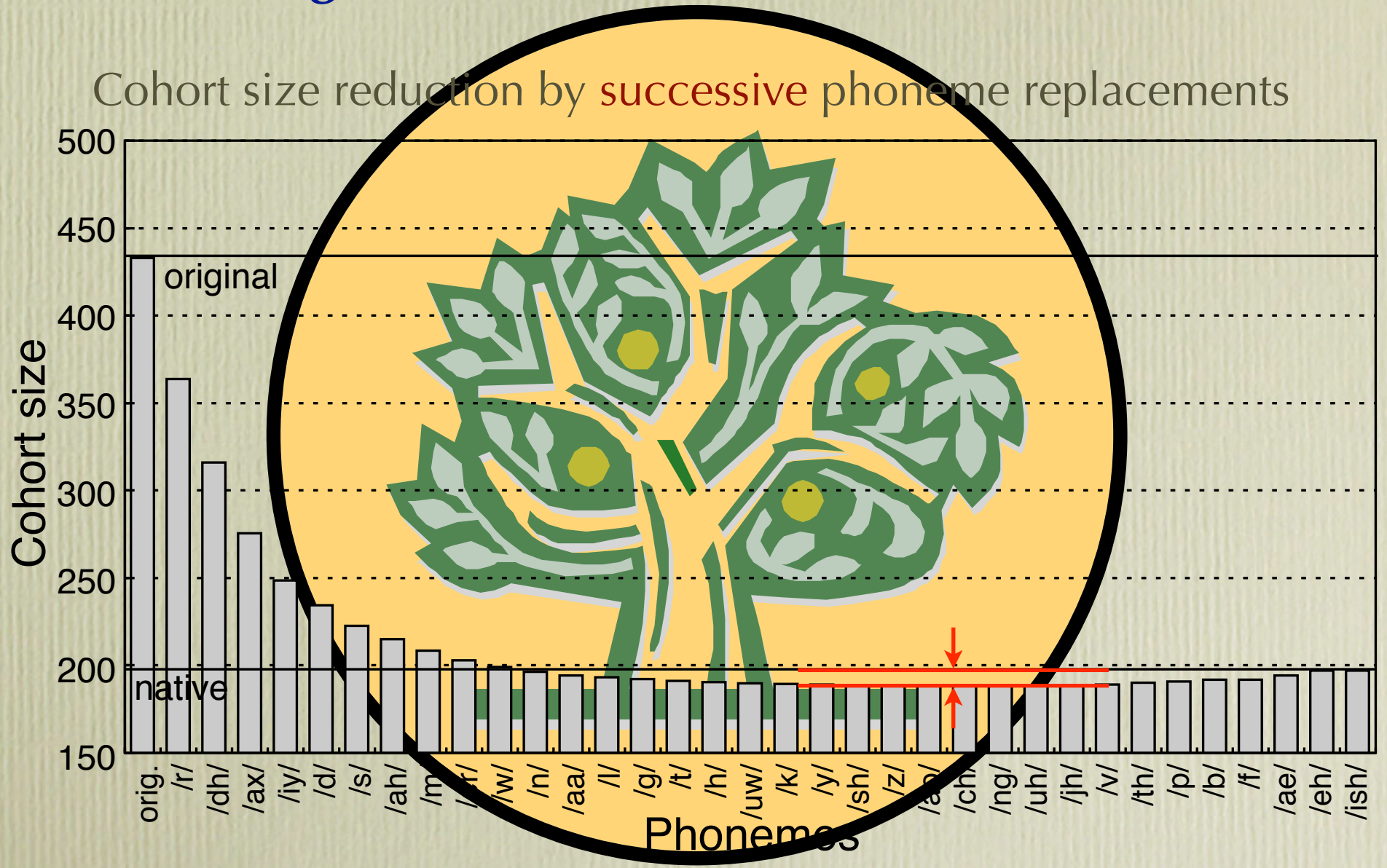
Possible applications of PTA (#3)

What is the order in the most efficient learning ?



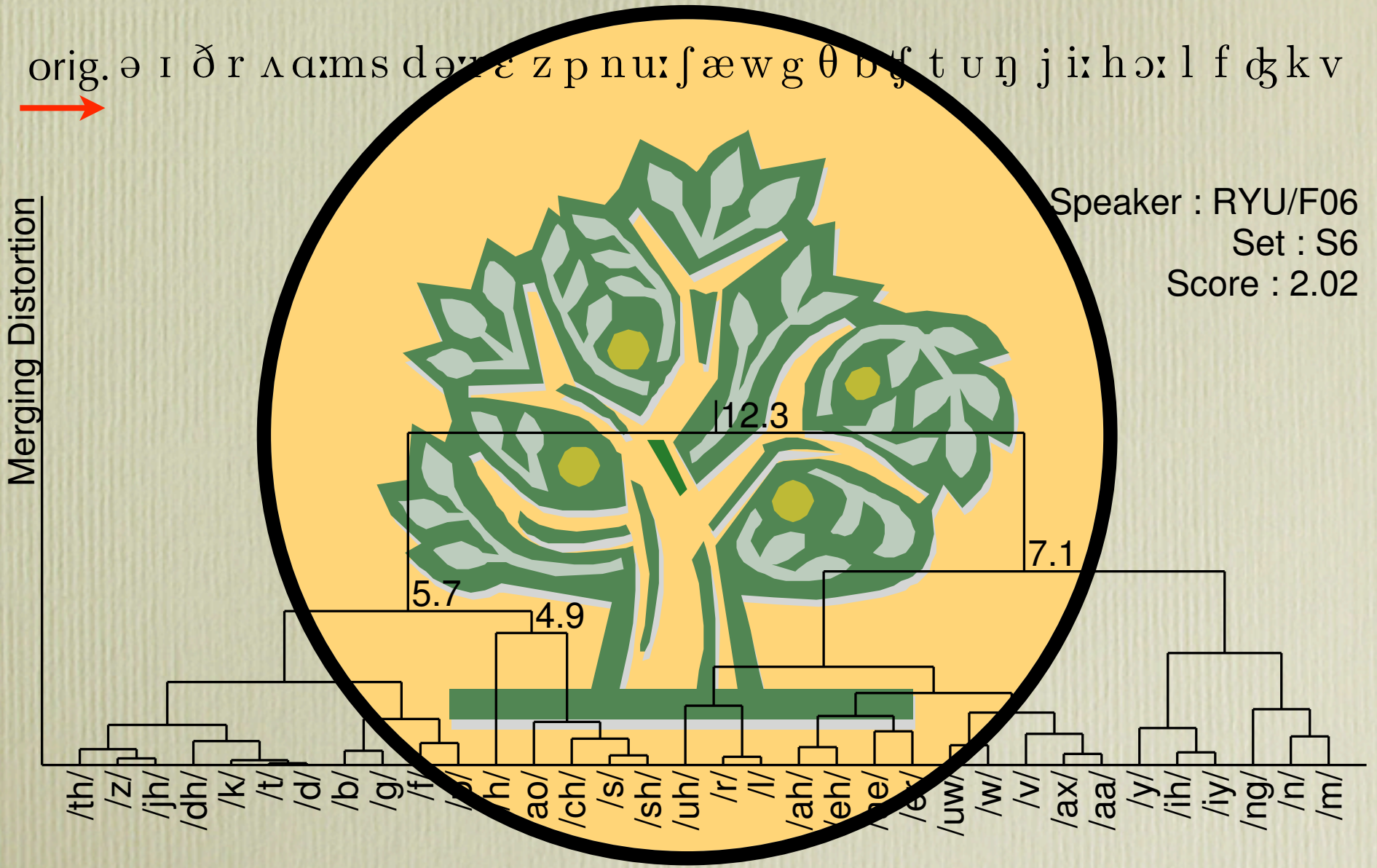
Possible applications of PTA (#4)

The learning order of another student



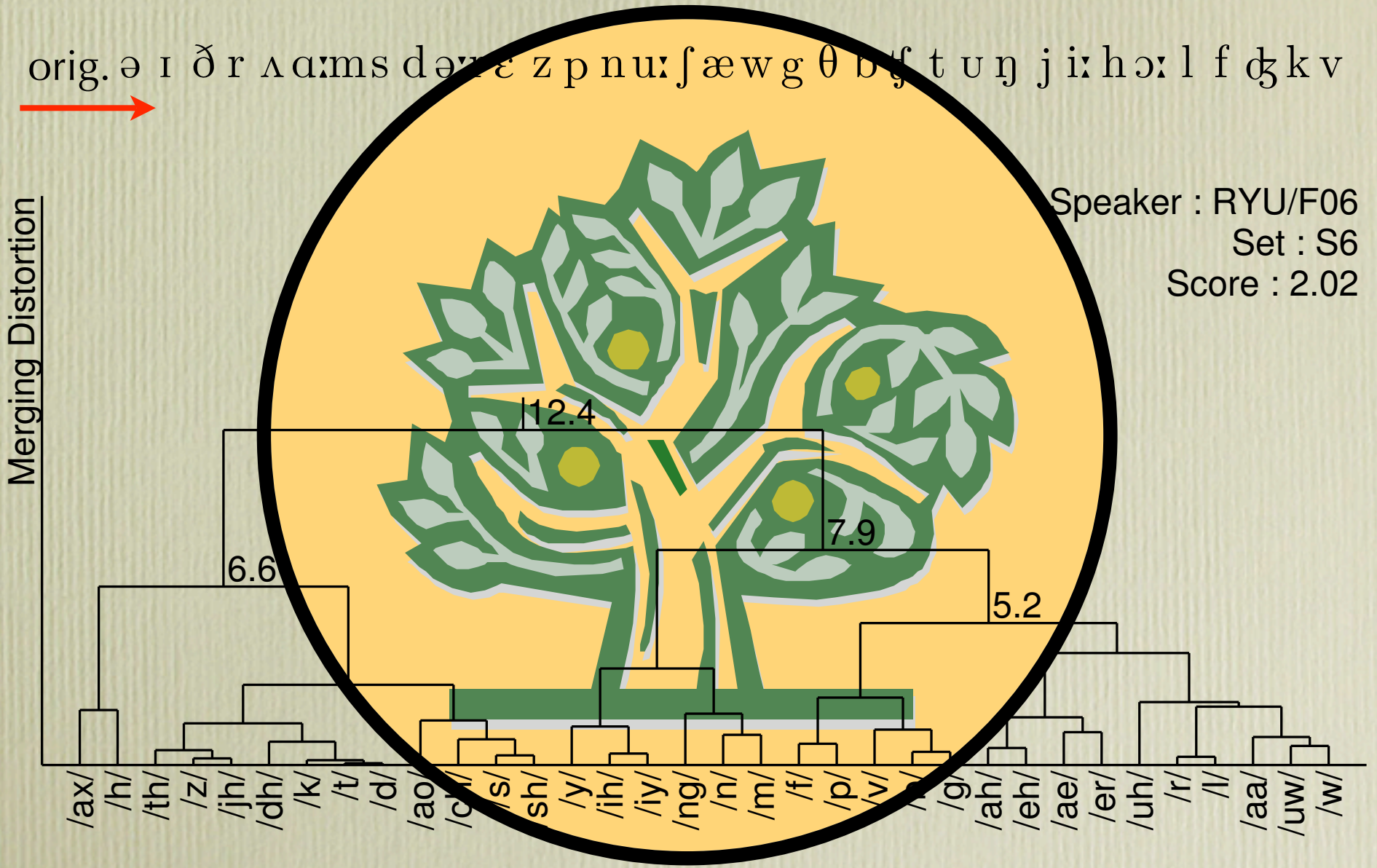
Possible applications of PTA (#5)

Prediction of her future.....?



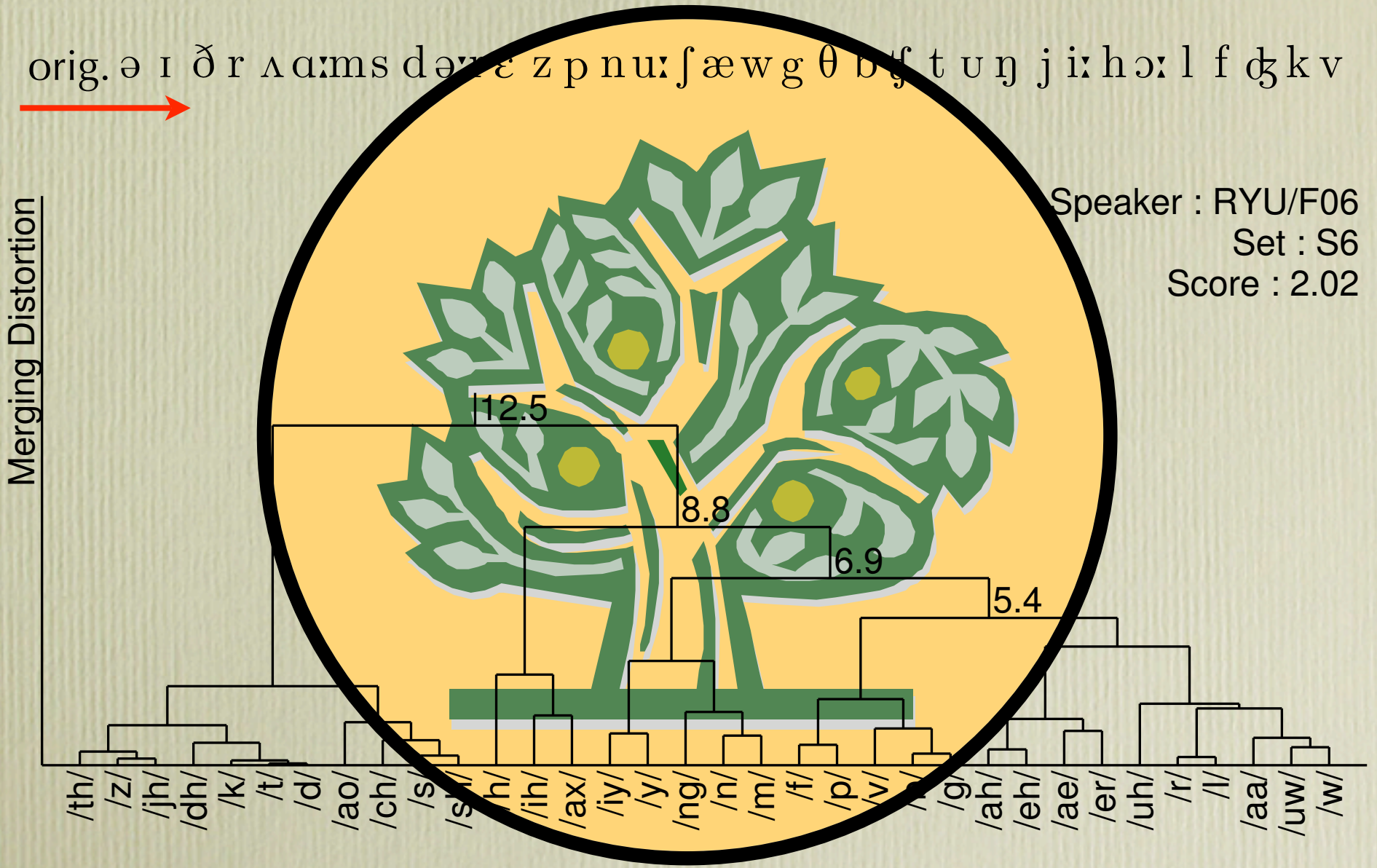
Possible applications of PTA (#5)

Prediction of her future.....?



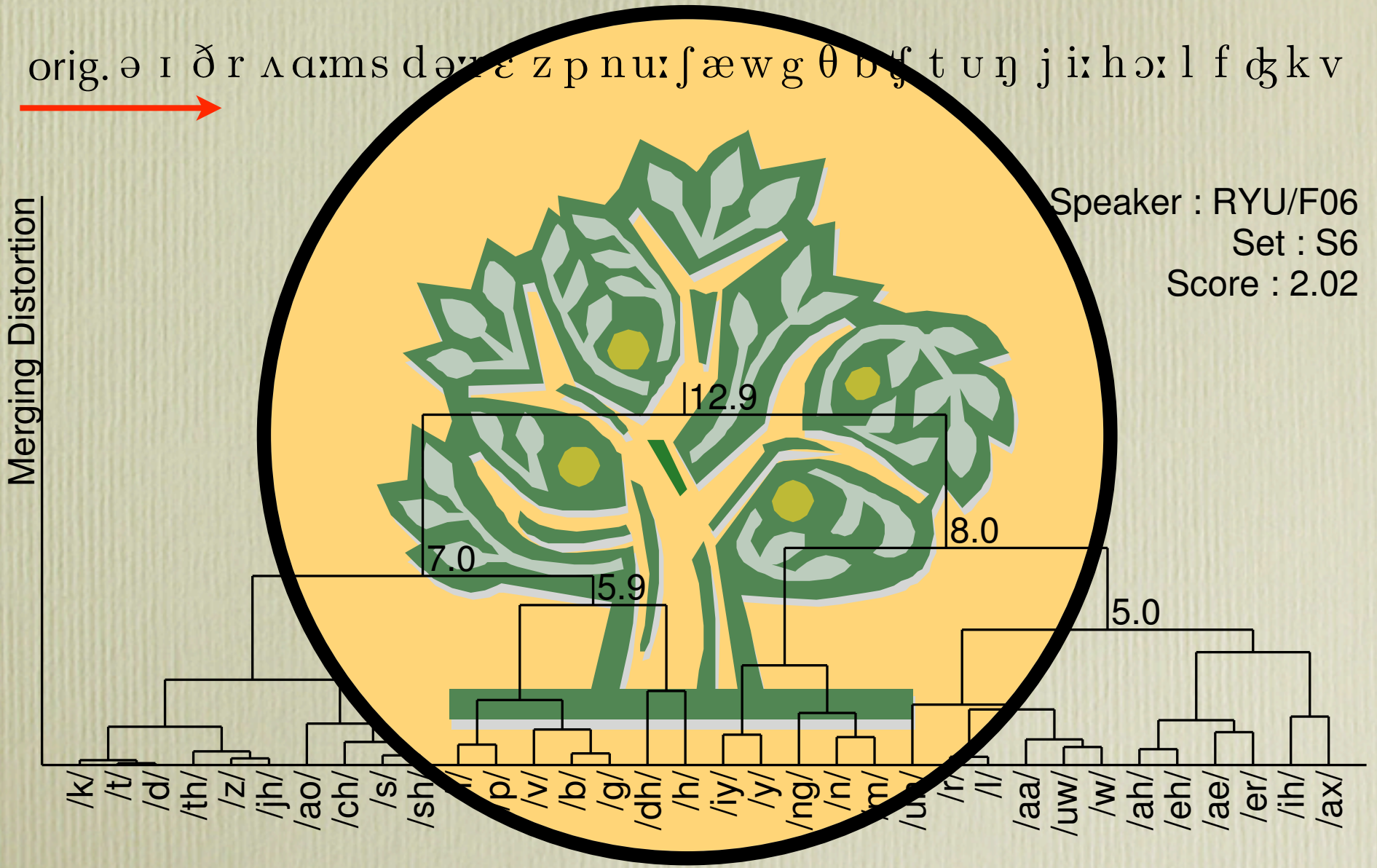
Possible applications of PTA (#5)

Prediction of her future.....?



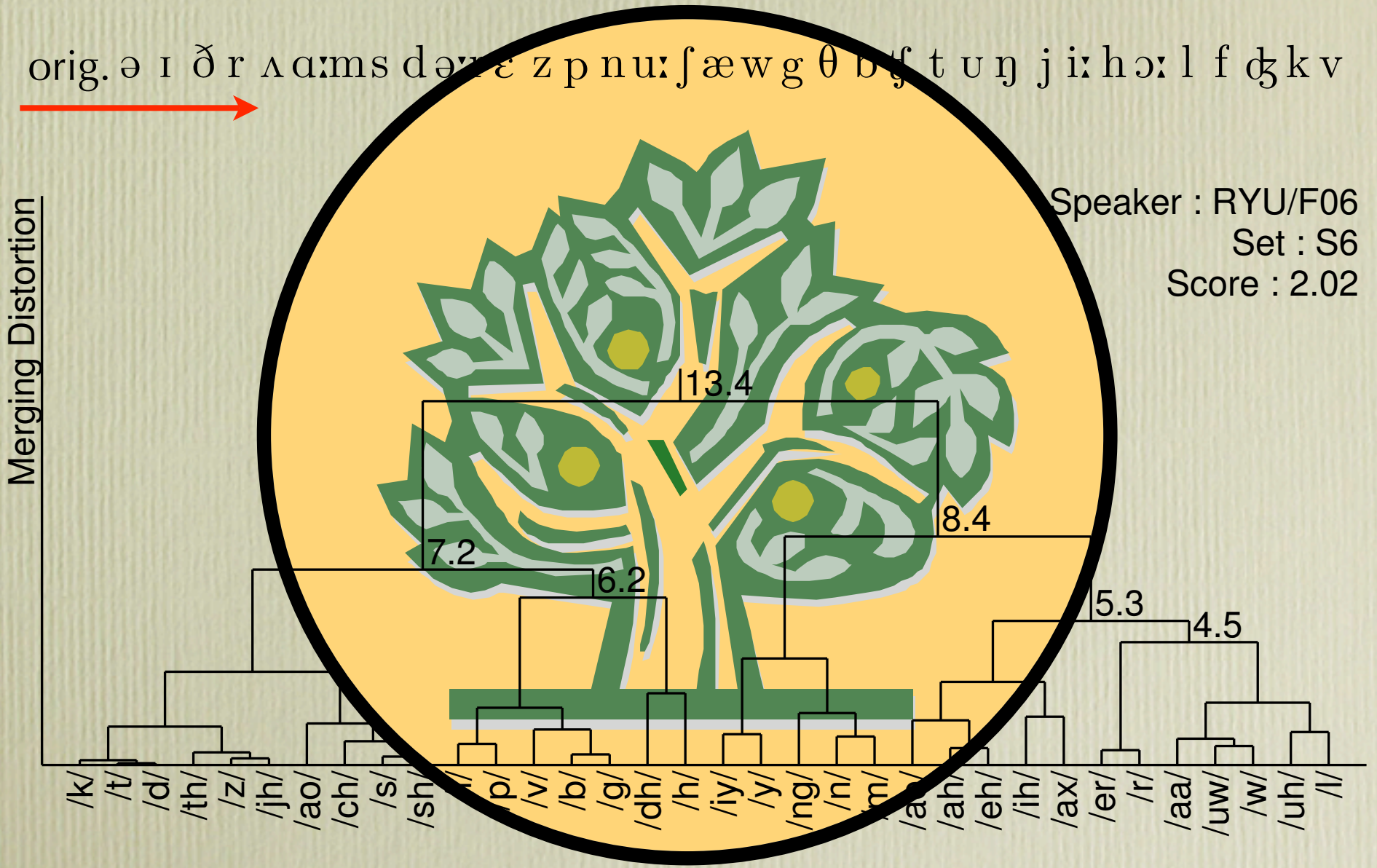
Possible applications of PTA (#5)

Prediction of her future.....?



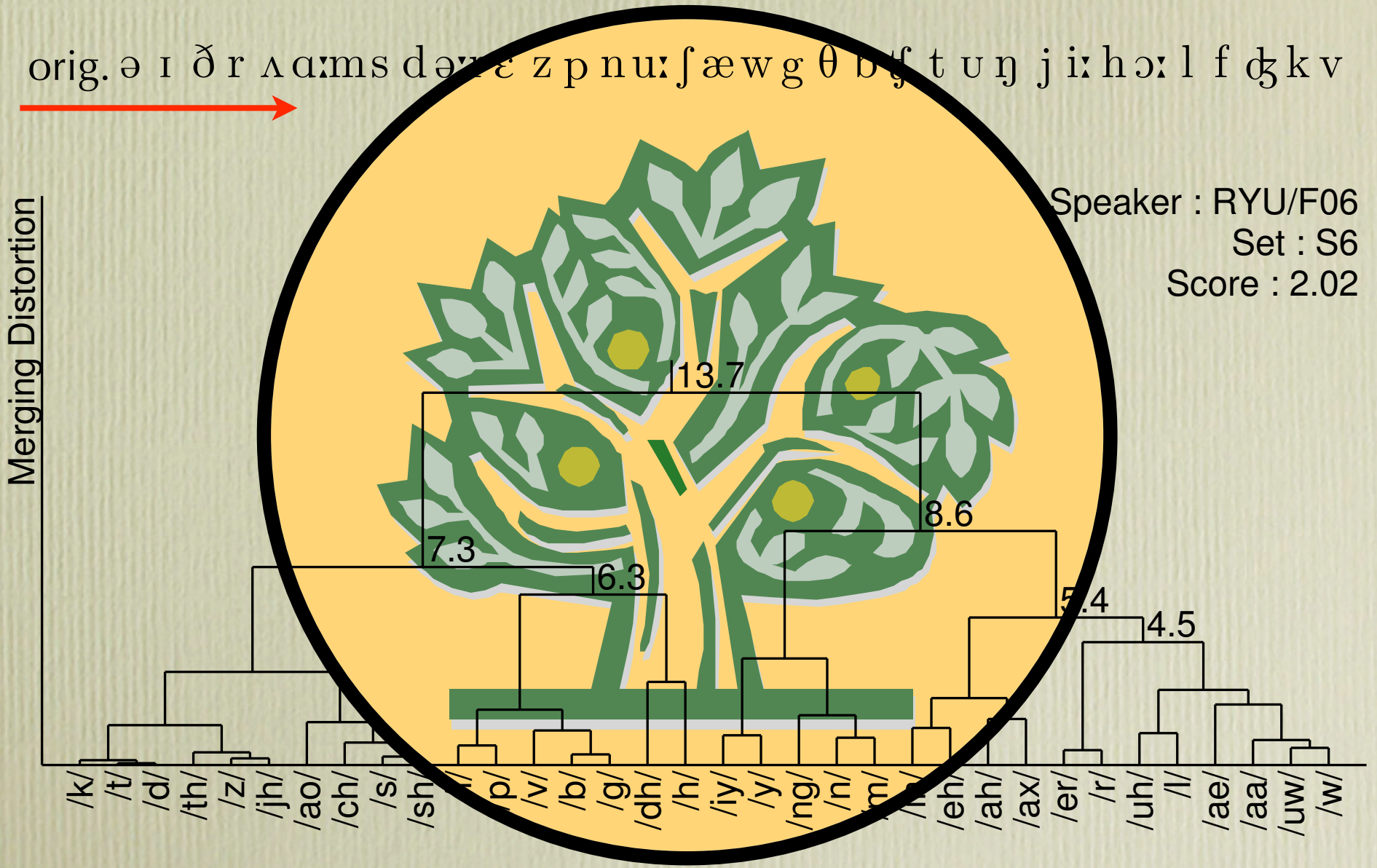
Possible applications of PTA (#5)

Prediction of her future.....?



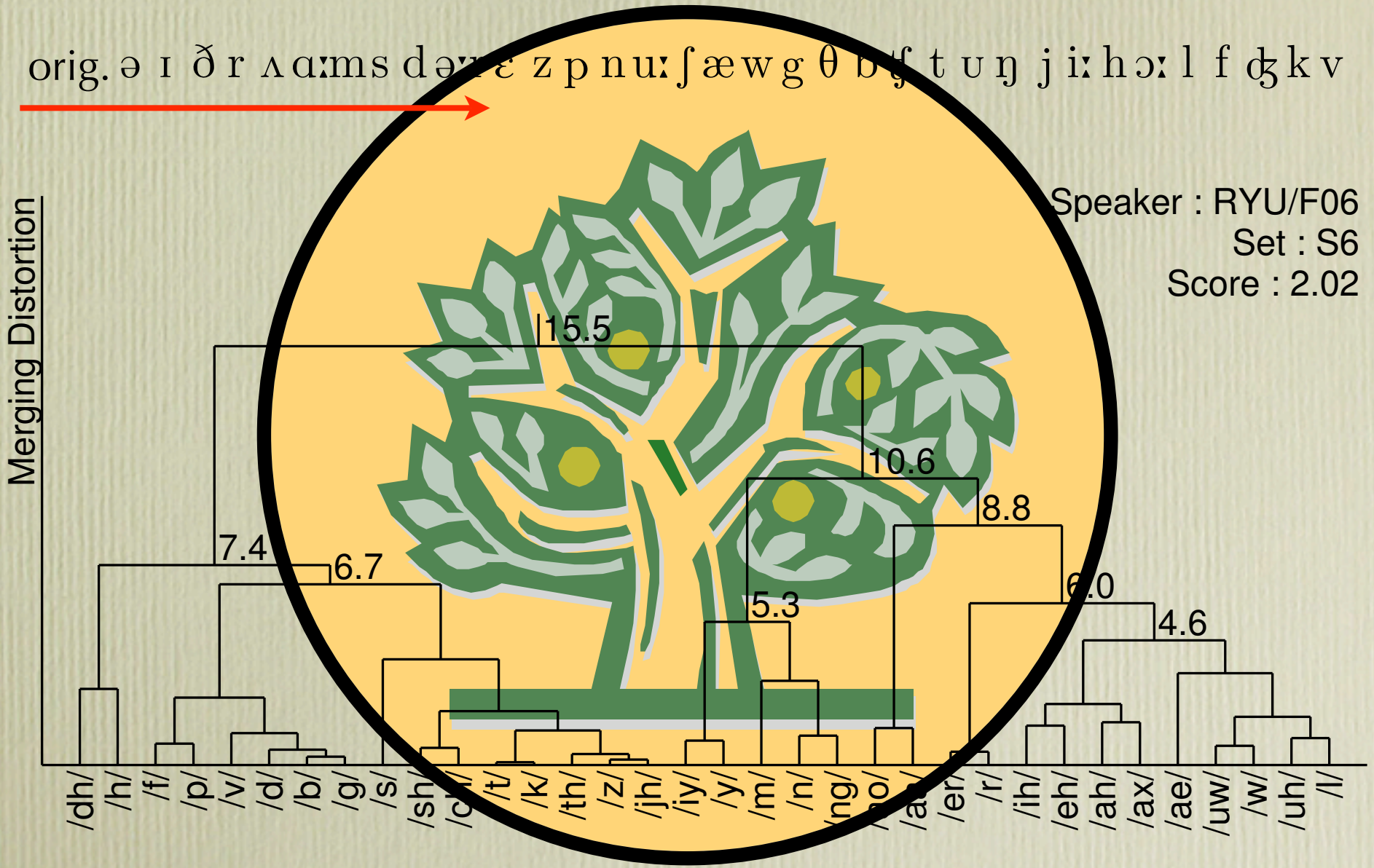
Possible applications of PTA (#5)

Prediction of her future.....?



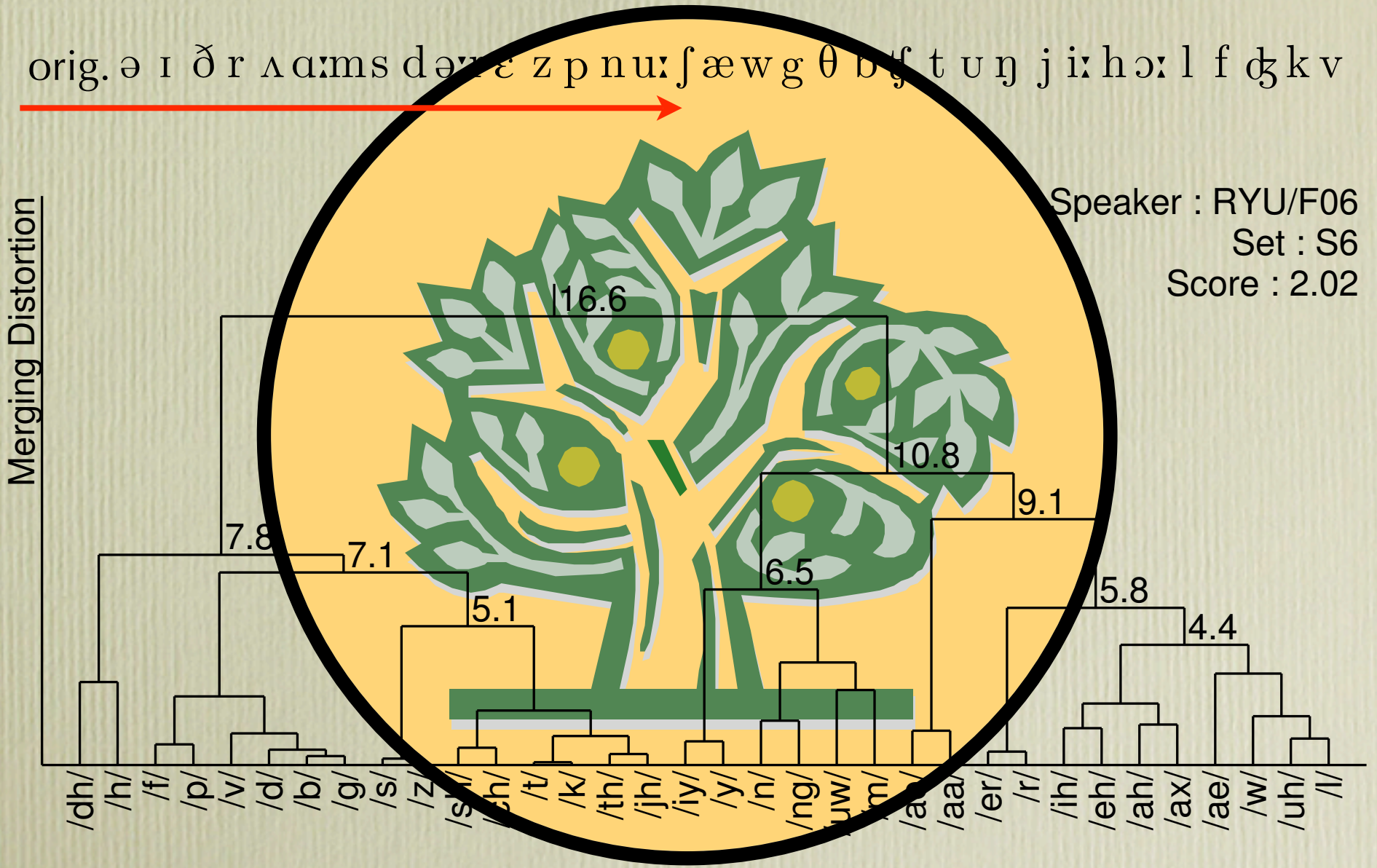
Possible applications of PTA (#5)

Prediction of her future.....?



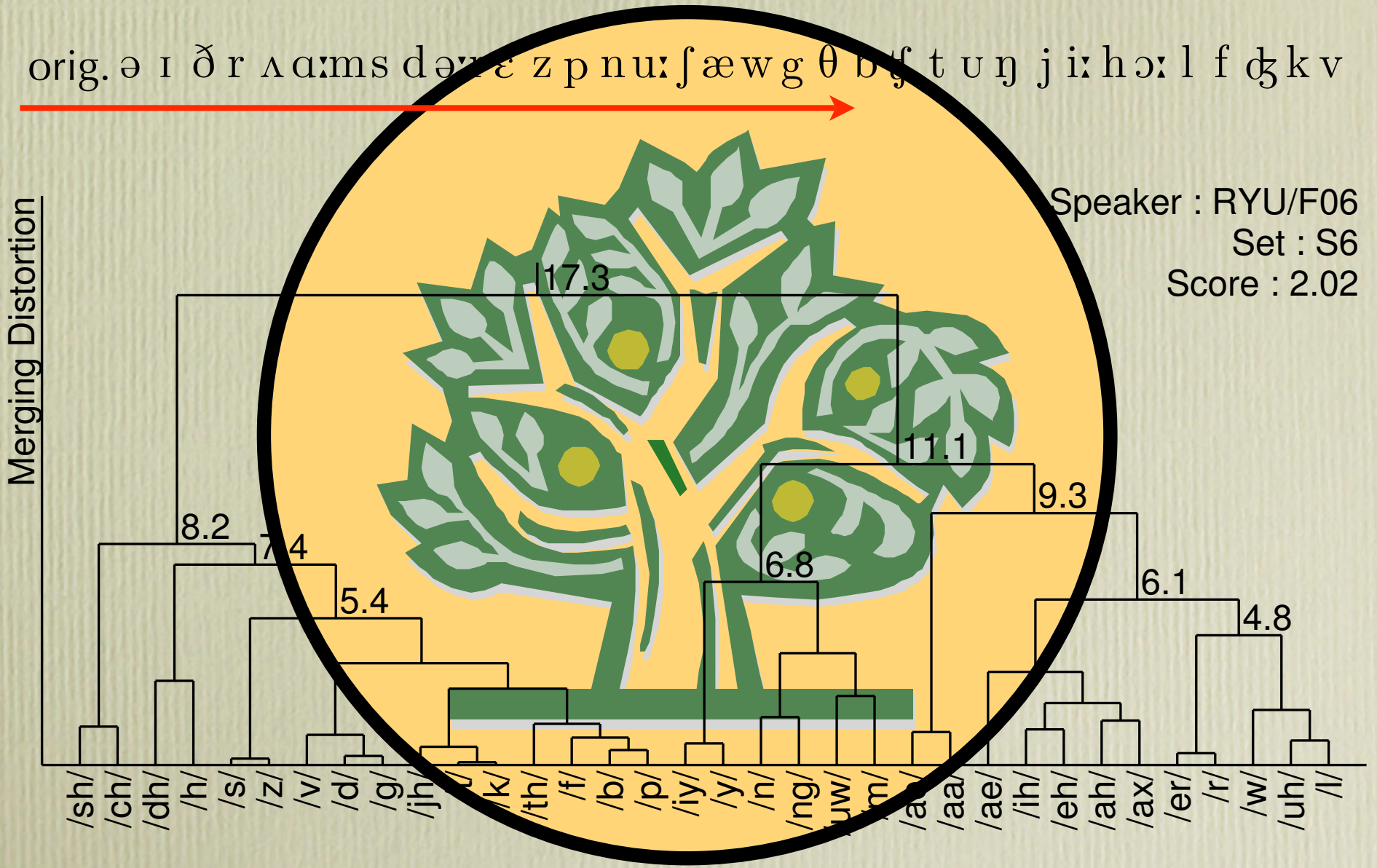
Possible applications of PTA (#5)

Prediction of her future.....?



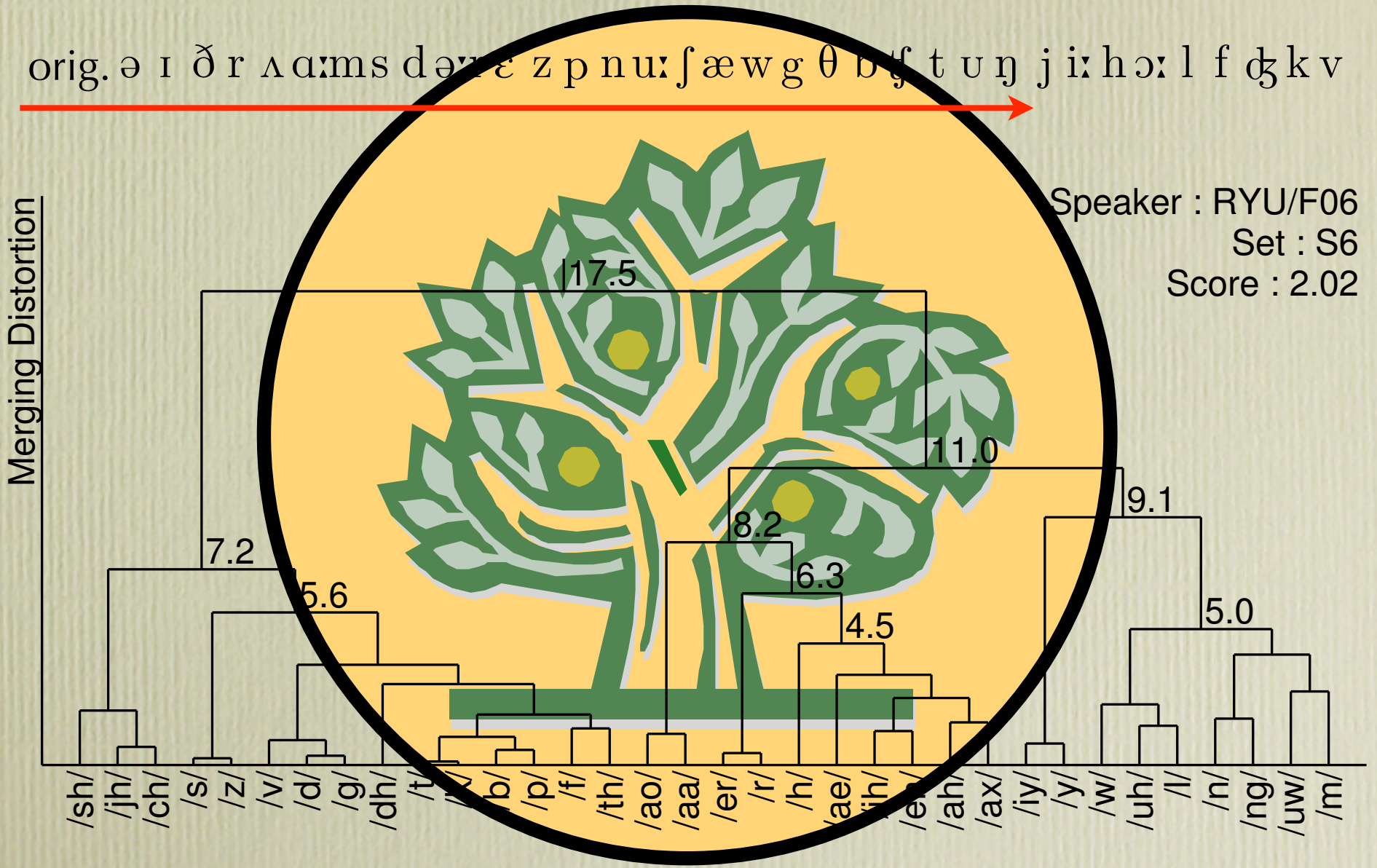
Possible applications of PTA (#5)

Prediction of her future.....?



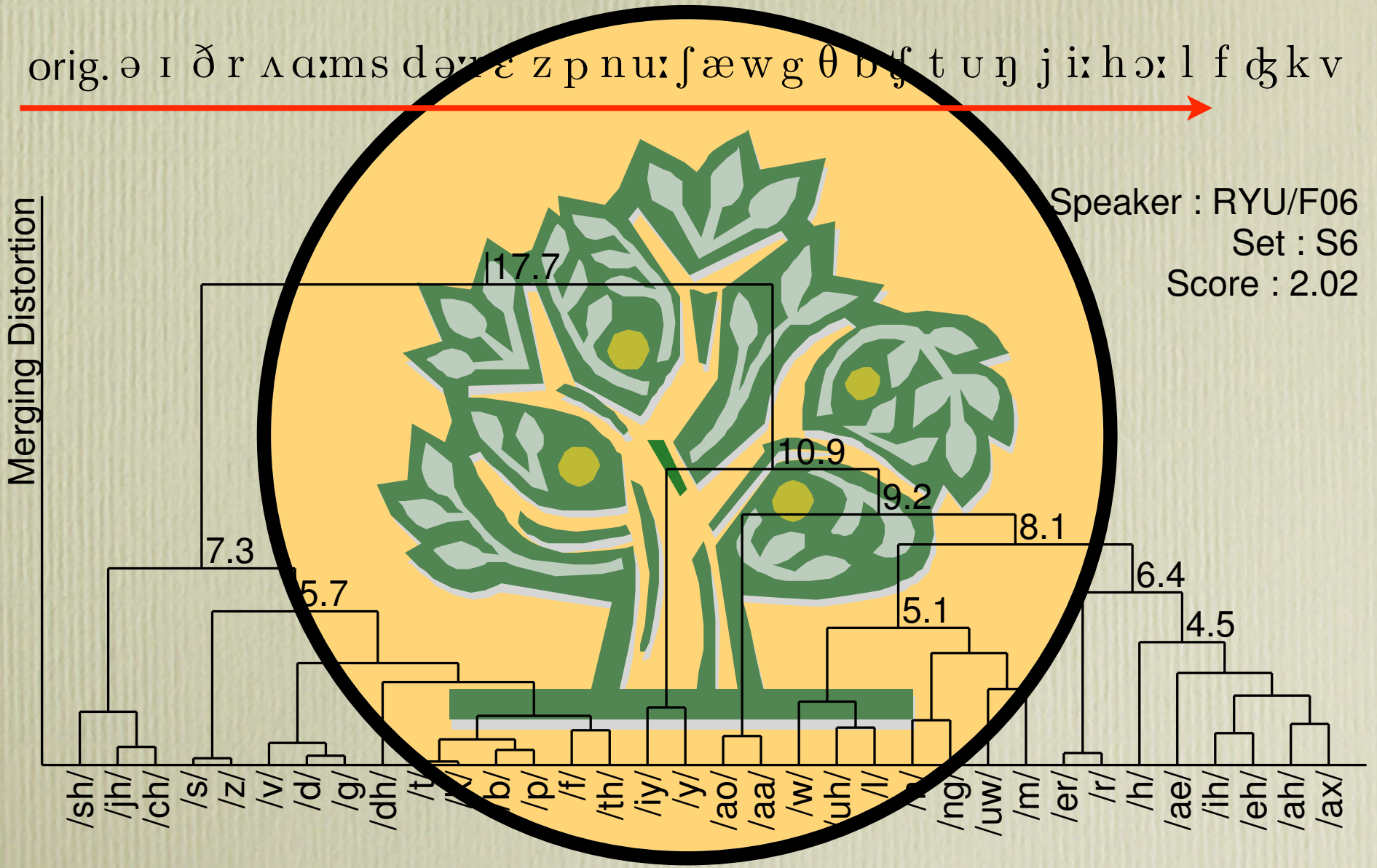
Possible applications of PTA (#5)

Prediction of her future.....?



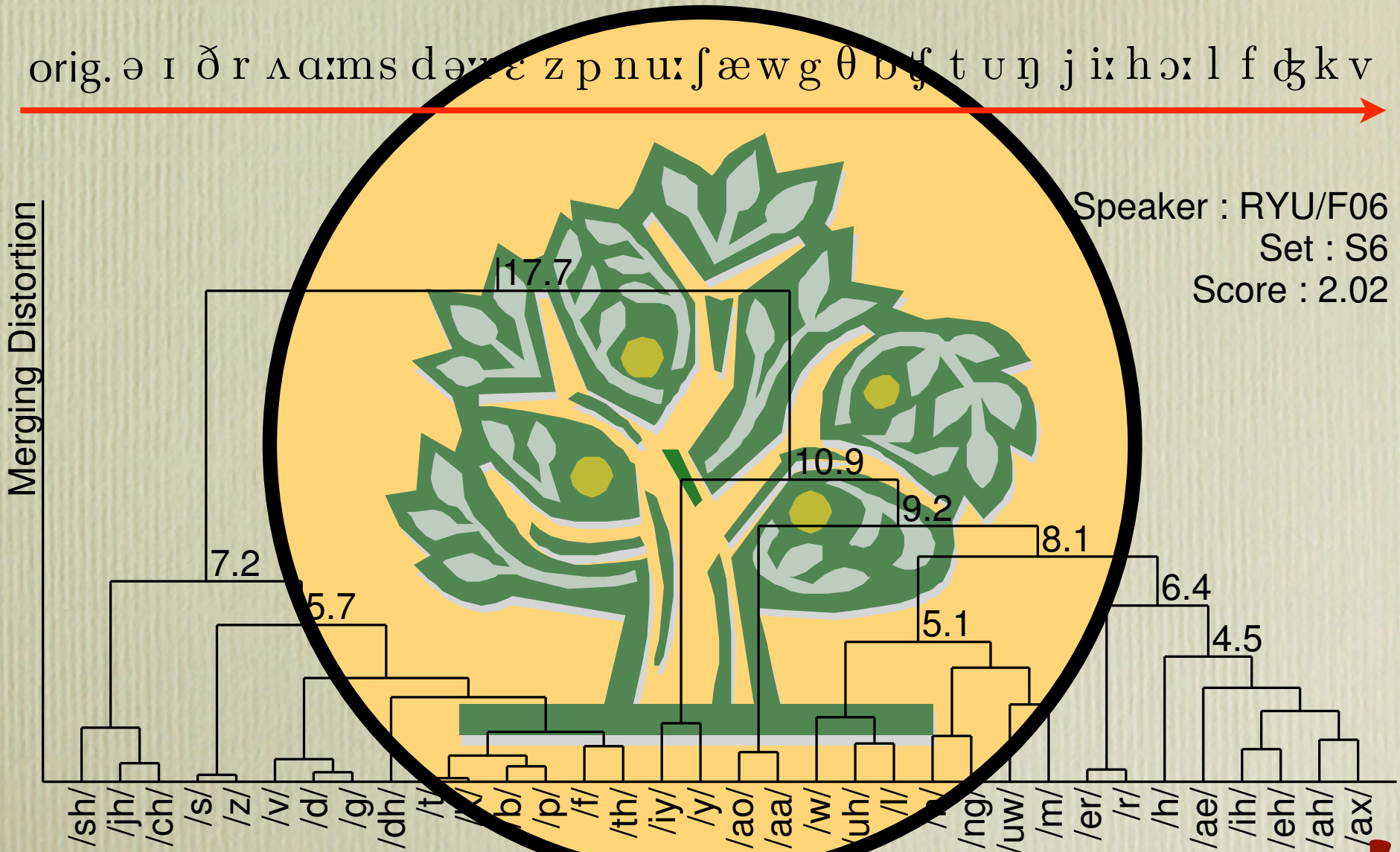
Possible applications of PTA (#5)

Prediction of her future.....?



Possible applications of PTA (#5)

Prediction of her future.....?



Goooooooooooooooooooooal !!

LVCSR vs. SIE

Two types of hearings

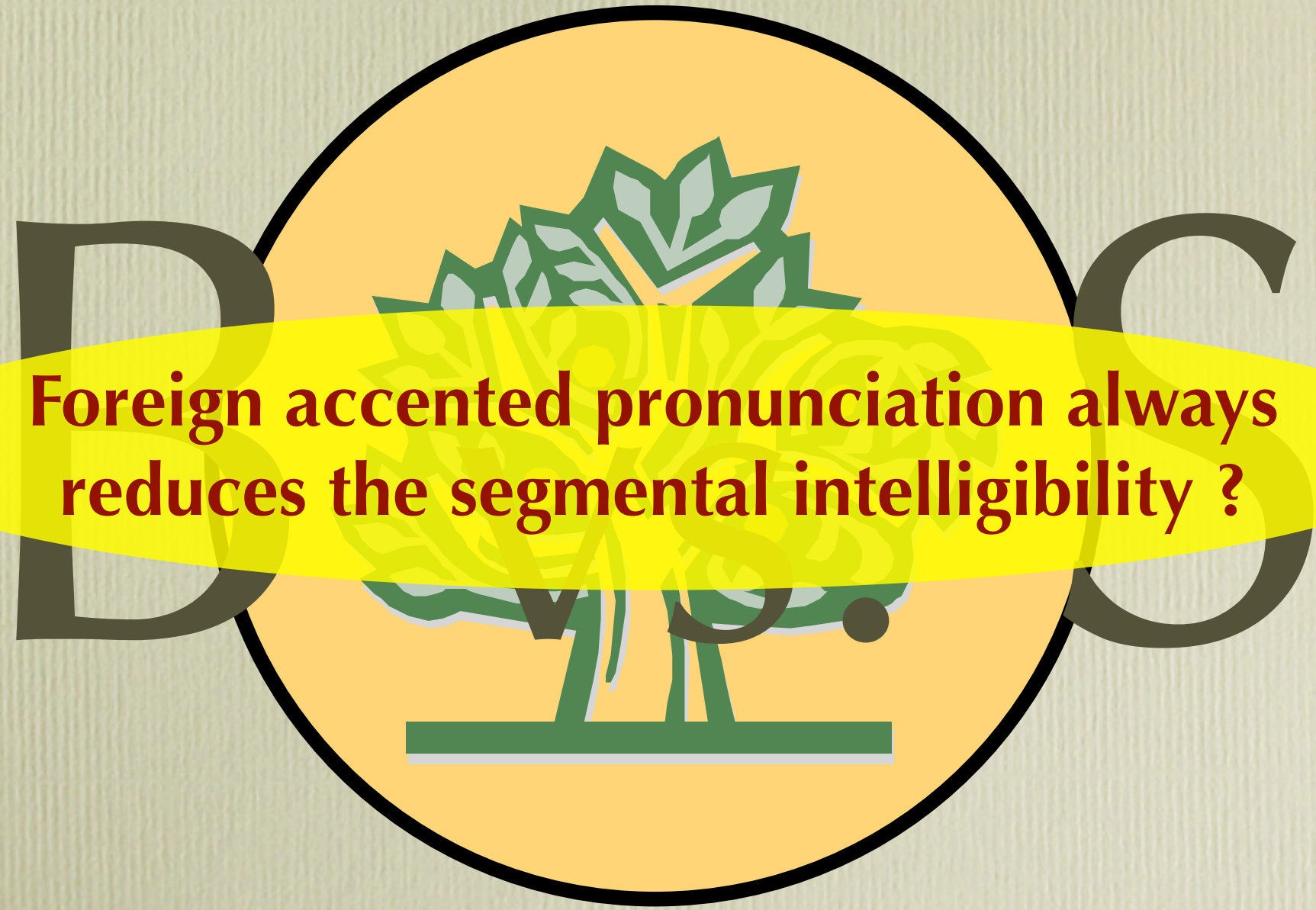
Input	an utterance (MFCC)	utterances (phonetic structure)
Acoustic model	phone models (tied-state triphones)	(native phonetic structure)
Pron. lexicon	phoneme-based tree structured lexicon	perceptual-unit-based tree structured lexicon
Lang. model	word trigram	word unigram (baseform unigram)
Integration	decoder	isolated word perception model
Output	sentence candidates	segmental intelligibility
Problems	mismatch children & elderly	~~mismatch children & elderly~~

native sounding

Some interesting issues on PTA (#1)

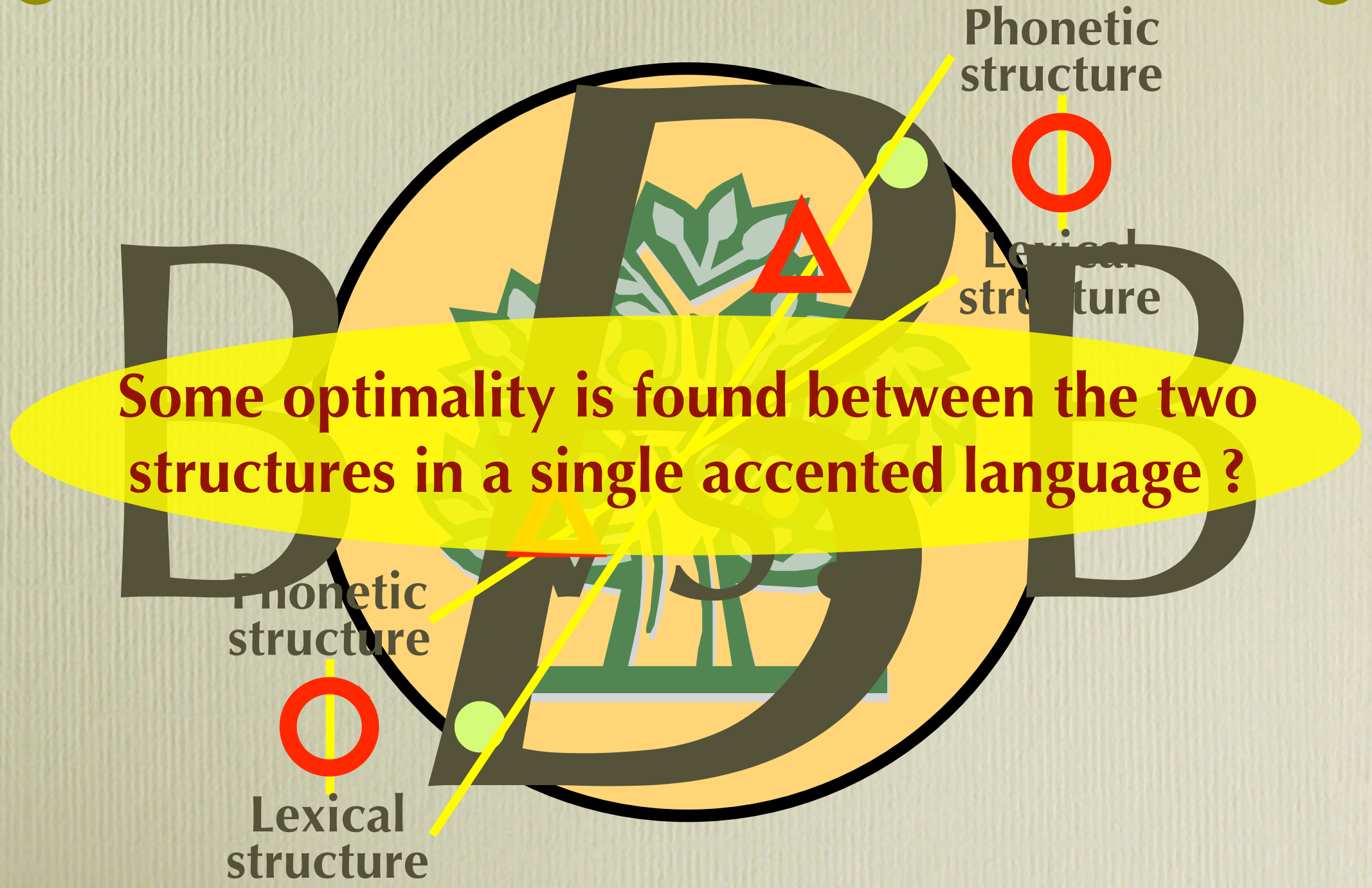
Which is more intelligible pronunciation ?

Some interesting issues on PTA (#2)

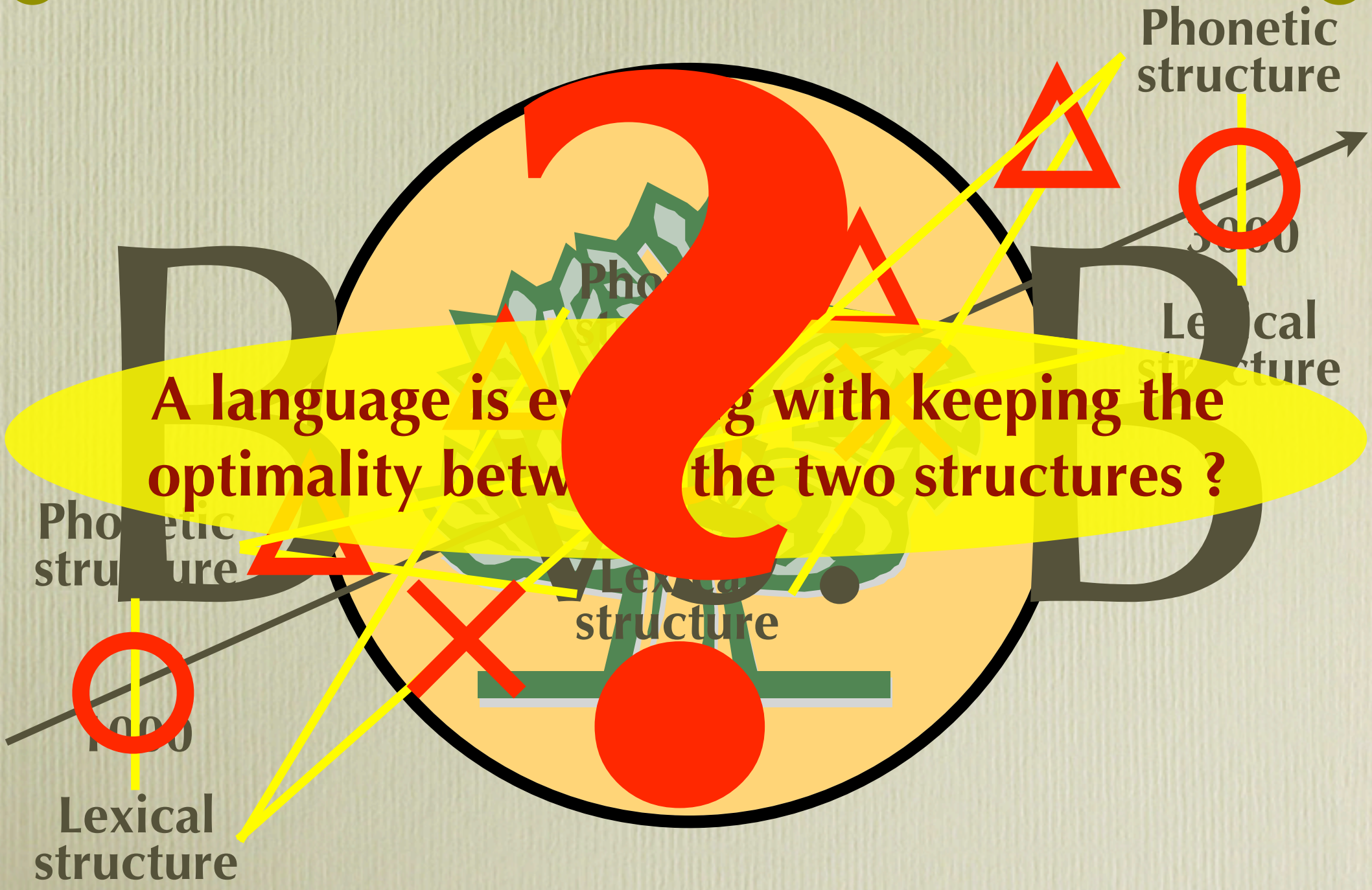


Foreign accented pronunciation always reduces the segmental intelligibility ?

Some interesting issues on PTA (#3)



Some interesting issues on PTA (#4)



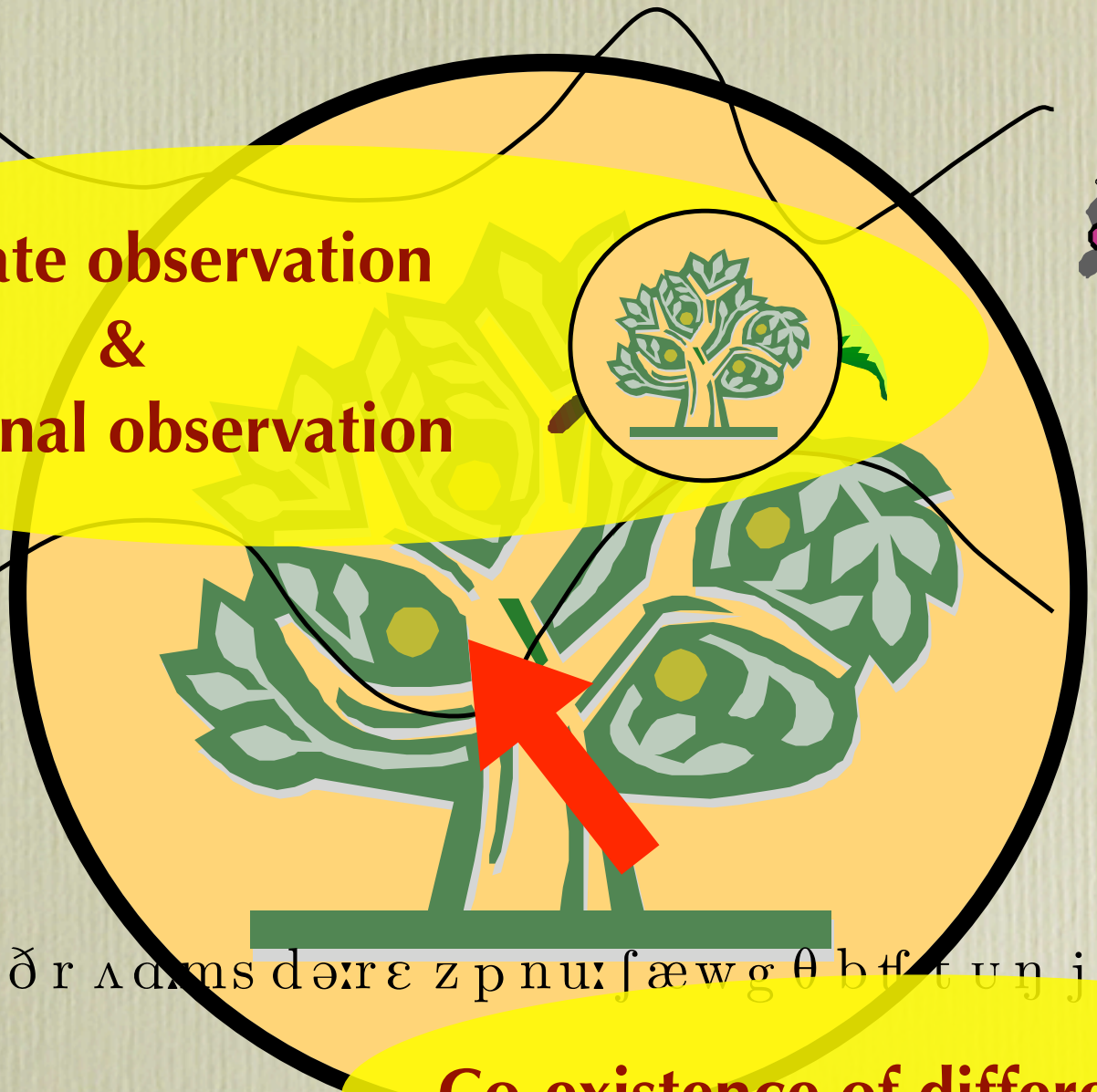
Two ways of looking at speech

Separate observation

&

Relational observation

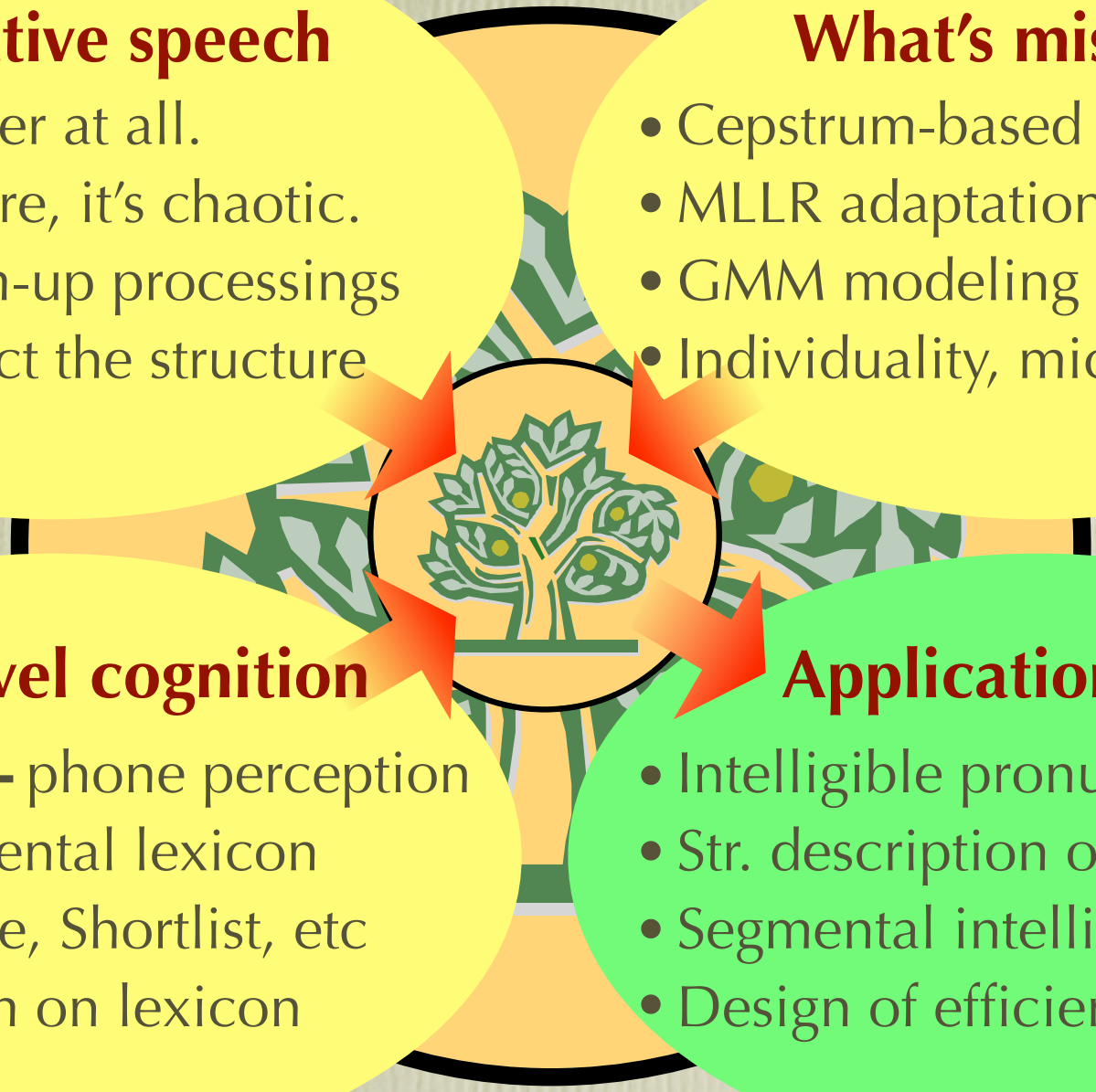
ɑː



ə ɪ ð r ʌ ɑː m s d əː r ɛ z p n uː ʃ æ w g θ b tʃ t ʊ ŋ j iː h ɔː l f ʤ k v

Co-existence of different phones

Several new lights on "Structuralism"



Non-native speech

- Learners differ at all.
- No rules there, it's chaotic.
- Only bottom-up processings
- PTA to extract the structure

What's missing ?

- Cepstrum-based space
- MLLR adaptation of ....
- GMM modeling of .....
- Individuality, mic, age, etc

Word-level cognition

- Phoneme ← phone perception
- Access to mental lexicon
- Cohort, Trace, Shortlist, etc
- Structuralism on lexicon

Application to LL

- Intelligible pronunciation
- Str. description of learners
- Segmental intelligibility
- Design of efficient learning

Conclusions & future works (#1)

Development of Japanese English read speech database

Corpus-based analysis of JE production

- Phonetic Tree Analysis (PTA)
- Relational observation can visualize well how the student is.

Corpus-based analysis of JE perception

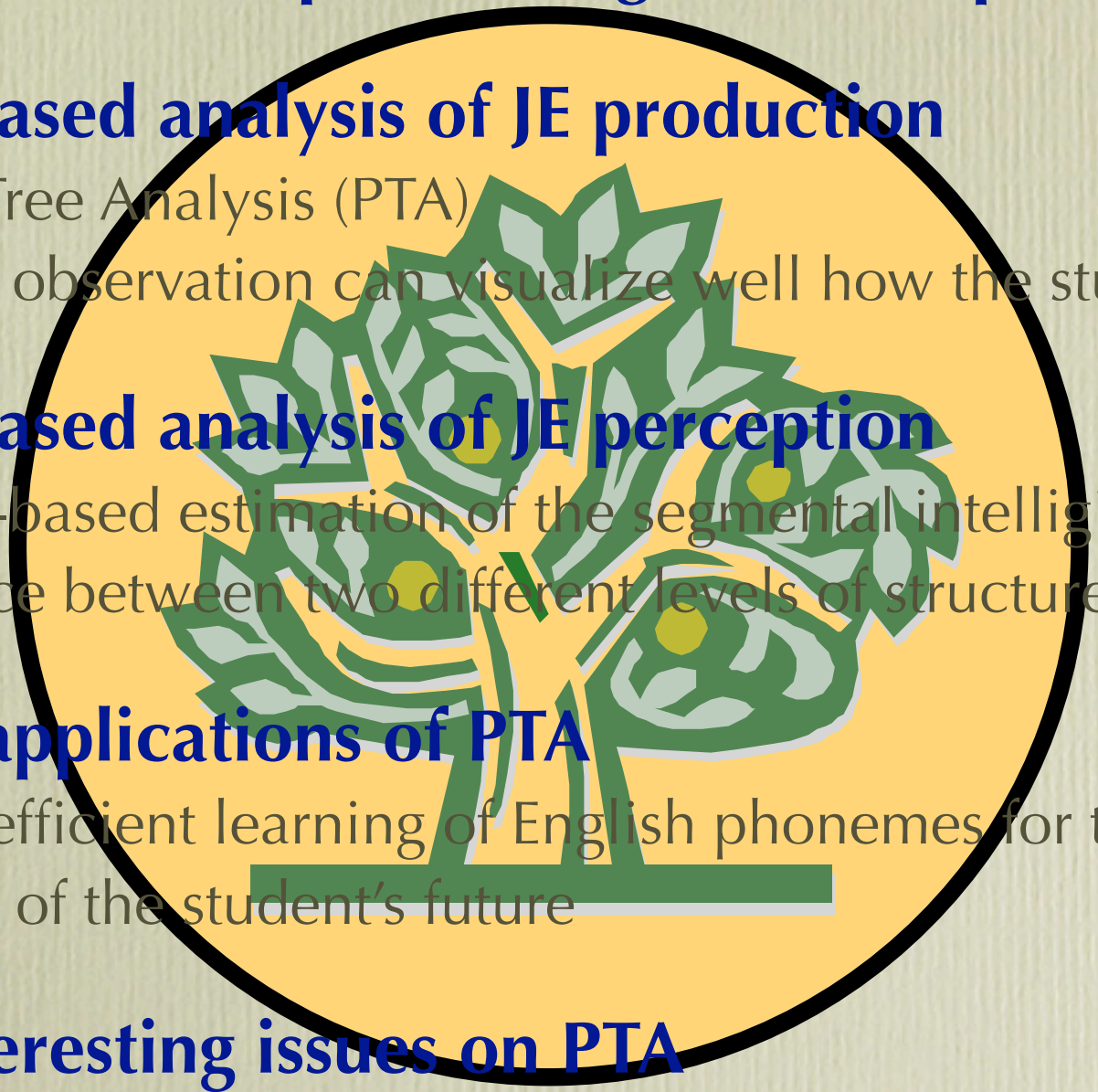
- Cognition-based estimation of the segmental intelligibility
- Accordance between two different levels of structures

Possible applications of PTA

- The most efficient learning of English phonemes for the student
- Prediction of the student's future

Some interesting issues on PTA

- B vs. A, B vs. S, and B vs. B



Conclusions & future works (#2)

Tuning up of acoustic conditions for analysis

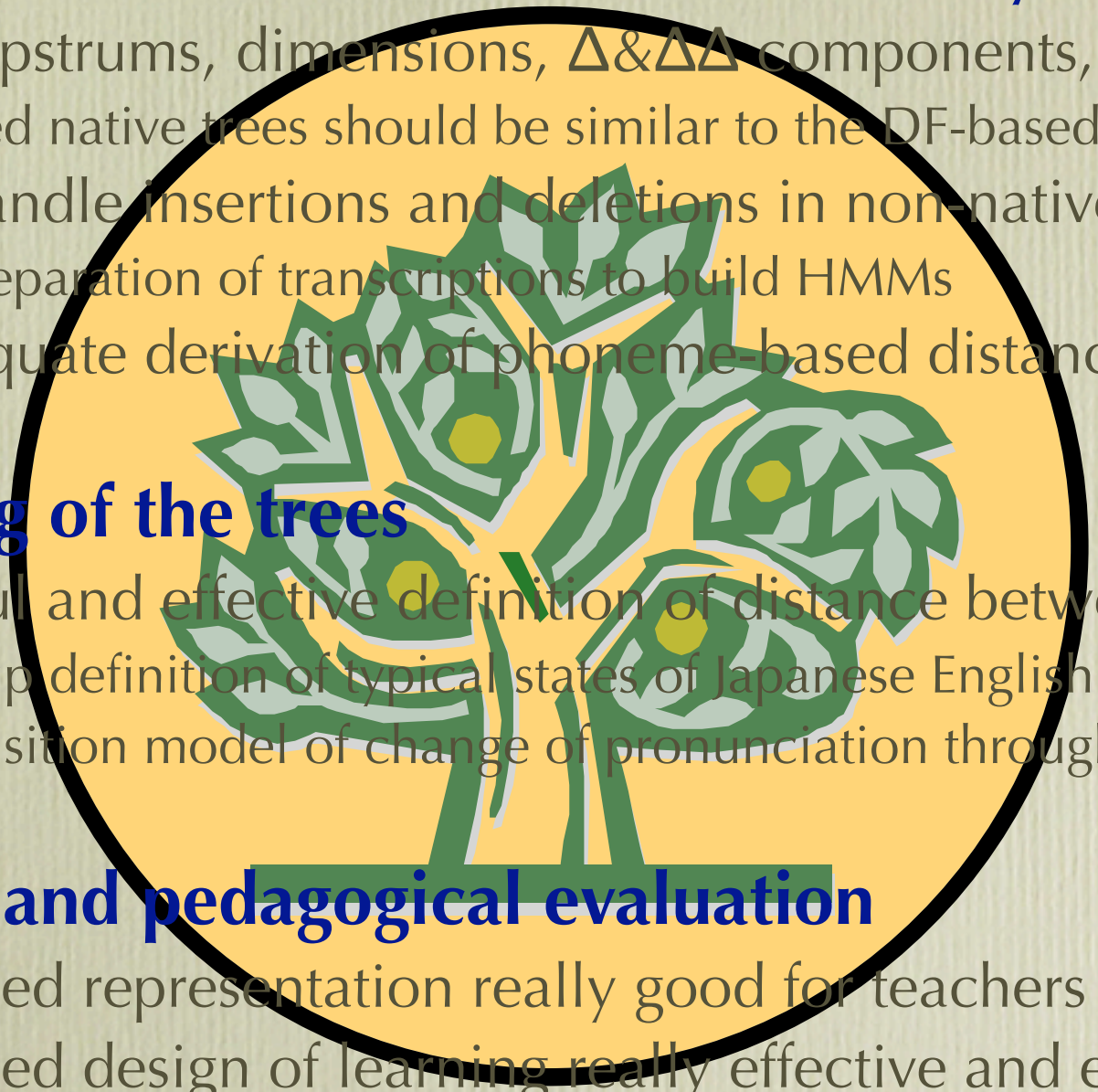
- Kind of cepstrums, dimensions, Δ&ΔΔ components, etc.
  - PTA-based native trees should be similar to the DF-based classical trees.
- How to handle insertions and deletions in non-native speech ?
  - Better preparation of transcriptions to build HMMs
- More adequate derivation of phoneme-based distance matrix

Clustering of the trees

- Meaningful and effective definition of distance between two trees
  - Bottom-up definition of typical states of Japanese English
  - State transition model of change of pronunciation through learning

Practical and pedagogical evaluation

- Is PTA-based representation really good for teachers and students ?
- Is PTA-based design of learning really effective and efficient ?




Thank you. Any questions ?