

# Computational Phonology, class 7: Modeling typology



Adam Albright

CreteLing 2022 — July 2022



[creteling2022.computational.phonology.party](https://creteling2022.computational.phonology.party)

# Universal CON?

- Assumption so far: the set of constraints is known ahead of time
  - Prince and Smolensky (2004): fixed and universal (UG)
- Assuming universal Con has been productive in reasoning about phonological typology
  - We'll consider the question of how one might learn constraints in the next class
- Factorial typology
  - Space of possible phonological grammars = permutations of ranked constraints
- How to assess factorial typology?
  - Empirical typology contains all and only the predicted patterns?
  - Overgeneration, undergeneration

# The role of learning

- Every attested adult grammar must be representable in the theory
- Not every representable grammar is necessarily expected to be attested
- Attested grammars must also be *acquirable*
  - Learners can arrive at it (priors+data)
- Predicted typology = representable languages, filtered through learning

- Undergeneration: fatal, if true (empirical adequacy)
  - But apparent exceptions merit careful scrutiny
- Overgeneration
  - Accidental gaps? (low expected probability, or historical 'accident')
  - Additional pressures, such as learnability



# Assigning stress without feet

- Constraints on position of stresses and intervals between them

ALIGN-L, ALIGN-R	Assign one * for each $\sigma$ separating stress from edge of the word
NONFINALITY	Assign one * for stress on the final $\sigma$
*LAPSE	Assign one * for each sequence of two stressless
*LAPSE-L/R	Assign one * if neither of the initial/final two stressed
*EXTENDLAPSE	Assign one * for each sequence of three stressless
*EXTLAPSE-L/R	Assign one * if none of the initial/final three stressed

# Assigning stress without feet (*cont.*)

- E.g., antepenultimate stress

/σσσσσσ/	*EXTLAPSE(R)	*EXTLAPSE(L)	ALIGN(L)	ALIGN(R)
a. όσσσσσσ	*! W		L	***** W
b. σόσσσσσ	*! W		* L	***** W
c. σσόςσσσ	*! W		** L	**** W
d. σσσόςσσ	*! W	*	*** L	*** W
☞ e. σσσσόςσ		*	****	**
f. σσσσσόςσ		*	*****! W	* L
g. σσσσσσός		*	*****!* W	L

- The insight behind the analysis
  - Stress wants to be as far left as possible (ALIGN(L)  $\gg$  ALIGN(R))
  - But it can't go further than the 3rd σ from the end (EXTLAPSE(R)  $\gg$  ALIGN(L))

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

- Languages may have stress at a fixed distance from an edge, but not all distances are observed

Attested:	final	ό, σό, σσό, σσσό, σσσσό, σσσσσό
Attested:	penultimate	ό, όσ, σός, σσός, σσσός, σσσσός
Attested:	antepenultimate	ό, όσ, όσσ, σόςσ, σσόςσ, σσσόςσ
Unattested:	preantepenultimate (and beyond)	ό, όσ, όσσ, όσσσ, σόςσσ, σσόςσσ



## Stress windows (cont.)

- Fixed CON solution: limit on constraints
  - \*EXTENDED LAPSE(R) penalizes όσσσ, prefers antepenultimate όσσ
  - Hypothesis: no equivalent \*EXTENDED EXTENDED LAPSE(R) (\*όσσσσ)

/σσσσσσ/	*EXT LAPSE(R)	*EXT LAPSE(L)	ALIGN(L)	ALIGN(R)
a. όσσσσσσ	*! W		L	***** W
b. σόσσσσσ	*! W		* L	***** W
c. σσόσσσσ	*! W		** L	**** W
d. σσσόσσσ	*! W	*	*** L	*** W
e. σσσσόσσ		*	****	**
f. σσσσσόσ		*	*****! W	* L
g. σσσσσσό		*	*****! * W	L

# Staub's general claim

- Phonological grammar allows for window lengths of any(?) size
- However, longer window lengths are harder to learn, because the data needed to distinguish them from short windows is rare
  - Long windows show up only in long words (following Prince 1993, Pater 2009)
- Harder to learn = lower probability that an individual learner will acquire it successfully
- Iterated across generations: frequency of such patterns is reduced or eliminated

# Illustrating the idea

- Staubs uses MaxEnt models with foot-based constraints, which intrinsically predict  $4+\sigma$  windows
- To keep things consistent, I'll recast the problem into the constraints already presented above, adding an additional constraint to allow for  $4\sigma$  windows
  - \*3-LAPSE(R/L): penalizes stressless  $\sigma\sigma\sigma\sigma\#$ ,  $\#\sigma\sigma\sigma\sigma$
  - With this notation,  $*\text{EXTLAPSE} = *2\text{-LAPSE}$

# Amount of ranking data

/σσ/	*3-LAPSE(L)	*3-LAPSE(R)	*2-LAPSE(L)	*2-LAPSE(R)	ALIGN-L	ALIGN-R
a. όσ						1
b. σό					1	
/σσσ/	*3-LAPSE(L)	*3-LAPSE(R)	*2-LAPSE(L)	*2-LAPSE(R)	ALIGN-L	ALIGN-R
a. όσσ						2
b. σόσ					1	1
c. σσό					2	
/σσσσ/	*3-LAPSE(L)	*3-LAPSE(R)	*2-LAPSE(L)	*2-LAPSE(R)	ALIGN-L	ALIGN-R
a. όσσσ				1		3
b. σόσσ					1	2
c. σσός					2	1
d. σσσο			1		3	
/σσσσσ/	*3-LAPSE(L)	*3-LAPSE(R)	*2-LAPSE(L)	*2-LAPSE(R)	ALIGN-L	ALIGN-R
a. όσσσσ		1		1		4
b. σόσσσ				1	1	3
c. σσόςσ					2	2
d. σσσοσ			1		3	1
e. σσσσό	1		1		4	



# Making the model sensitive to amount of data

- Gradual learning: fewer relevant W/L pairs means fewer updates
- Priors/regularization: less data means the prior gets more of a say

# Why would priors favor shorter windows?

- Given ambiguous data, why would learners favor shorter windows?
  - Train on  $4\sigma$  window, but mislearn as  $3\sigma$  window?
- $4\sigma$  window:  $*_3\text{-LAPSE}(R) \gg \text{ALIGN-L} \gg *_2\text{-LAPSE}(R)$
- $3\sigma$  window:  $*_2\text{-LAPSE}(R) \gg \text{ALIGN-L}$ 
  - Ranking of  $*_2\text{-LAPSE}(R)$  doesn't matter
- Idea: if constraints are promoted,  $*_2\text{-LAPSE}(R)$  has more data favoring its promotion
- Even if data has words of all lengths, more inputs promote  $*_2\text{-LAPSE}(R)$ 
  - $4\sigma$  and up for  $*_2\text{-LAPSE}$ , vs.  $5\sigma$  and up for  $*_3\text{-LAPSE}$

# Midpoint Pathology



# The midpoint pathology (Kager, 2012; Stanton, 2016)

- Short words: can satisfy both  $*(\text{EXTENDED})\text{LAPSE}(L)$  and  $*(\text{EXTENDED})\text{LAPSE}(R)$ , by keeping stress towards the middle of the word

/σσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)
☞ a. σσόςσ		
b. σσσός	*! W	
c. όσσσσ		*! W

- Longer words: can't satisfy both, so satisfy the higher-ranked one with stress at the relevant edge

/σσσσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)
a. σσσόςσσ	*! W	*
☞ b. όσσσσσσ		*
c. σσσσσσός	*! W	L



## Example: a 'midpoint-stress' language

$*\text{EXTLAPSE}(\text{L}) \gg * \text{EXTLAPSE}(\text{R}) \gg \text{ALIGN}(\text{L}) \gg \text{ALIGN}(\text{R})$

2 syl     $\acute{\sigma}$

3 syl     $\acute{\sigma}\sigma$

4 syl     $\sigma\acute{\sigma}\sigma$

5 syl     $\sigma\sigma\acute{\sigma}\sigma$

6 syl     $\acute{\sigma}\sigma\sigma\sigma\sigma$

7 syl     $\acute{\sigma}\sigma\sigma\sigma\sigma\sigma$

8 syl     $\acute{\sigma}\sigma\sigma\sigma\sigma\sigma\sigma$

- $*\text{EXTLAPSE}(\text{L/R}) \gg \text{ALIGN}(\text{L/R})$ : stress can move inside word to avoid extended lapse
- $*\text{EXTLAPSE}(\text{L}) \gg * \text{EXTLAPSE}(\text{R})$ : when too long to satisfy both, it moves to the left side of the word
- $\text{ALIGN}(\text{L}) \gg \text{ALIGN}(\text{R})$ : when on the left side of the word, it falls on the very first syllable

# Stanton's observation

/σσ/	*ExtLAPSE(L)	*ExtLAPSE(R)	ALIGN(L)	ALIGN(R)
ΕΠ a. όσ			*! W	*
b. σό				L
/σσσ/	*ExtLAPSE(L)	*ExtLAPSE(R)	ALIGN(L)	ALIGN(R)
ΕΠ a. όσσ			*! W	**
b. σόσ			*! W	* L
c. σσά			*! W	L
/σσσσ/	*ExtLAPSE(L)	*ExtLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσ		*! W	L	*** W
ΕΠ b. σόσσ			*	**
c. σσάσ			**! W	* L
d. σσσά	*! W		*** W	L
/σσσσσ/	*ExtLAPSE(L)	*ExtLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσσ		*! W	L	**** W
b. σόσσσ		*! W	* L	*** W
ΕΠ c. σσάσσ			**	**
d. σσσάσ	*! W		*** W	* L
e. σσσσά	*! W		**** W	L
/σσσσσσ/	*ExtLAPSE(L)	*ExtLAPSE(R)	ALIGN(L)	ALIGN(R)
ΕΠ a. όσσσσσ		*		*****
b. σόσσσσ		*	* W	**** L
c. σσάσσσ		*	** W	*** L
d. σσσάσσ	*! W	L	*** W	** L
e. σσσσάσ	*! W	L	**** W	* L
f. σσσσσά	*! W	L	***** W	L
/σσσσσσσ/	*ExtLAPSE(L)	*ExtLAPSE(R)	ALIGN(L)	ALIGN(R)
ΕΠ a. όσσσσσσ		*		*****
b. σόσσσσσ		*	* W	***** L
c. σσάσσσσ		*	** W	**** L
d. σσσάσσσ	*! W	*	*** W	*** L
e. σσσσάσσ	*! W	L	**** W	** L
f. σσσσσάσ	*! W	L	***** W	* L

- Clear evidence for  $ALIGN(L) \gg ALIGN(R)$  in 2,3,4-syllable words
- Evidence for  $*EXTLAPSE(R) \gg ALIGN(L)$  from 5-syllable words
- Evidence for  $*EXTLAPSE(L) \gg *EXTLAPSE(R)$  only from 6-syllable words and longer

# On the relative scarcity of long words

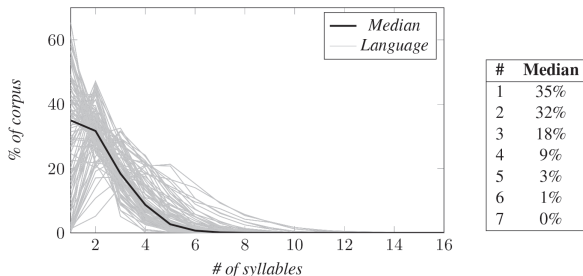


FIGURE 1. Results of the survey of text corpora from 102 languages (see the appendices for more details).

- Rough estimate of relative proportion of words of different lengths in texts of 102 languages
- With a few notable exceptions,  $\geq 6\sigma$  words are a very small proportion of the input
- Also: long words tend to be morphologically complex (may show other patterns)

# Learning from short words

/σσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☞ a. όσ				*
b. σό			* W	L

/σσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☞ a. όσσ				**
b. σόσ			* W	* L
c. σσό			** W	L

/σσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσ		* W	L	*** W
☞ b. σόσσ			*	**
c. σσός			** W	* L
d. σσσό	* W		*** W	L

/σσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσσ		* W	L	**** W
b. σόσσσ		* W	* L	*** W
☞ c. σσόςσ			**	**
d. σσσός	* W		*** W	* L
e. σσσσό	* W		**** W	L

# Learning from short words

/σσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☞ a. όσ				*
b. σό			* W	L

/σσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☞ a. όσσ				**
b. σόσ			* W	* L
c. σσό			** W	L

/σσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσ		* W	L	*** W
☞ b. σόσσ			*	**
c. σσός			** W	* L
d. σσσό	* W		*** W	L

/σσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσσ		* W	L	**** W
b. σόσσσ		* W	* L	*** W
☞ c. σσόςσ			**	**
d. σσσός	* W		*** W	* L
e. σσσσό	* W		**** W	L

# Learning from short words

/σσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☞ a. όσ				*
b. σό			* W	L

/σσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☞ a. όσσ				**
b. σόσ			* W	* L
c. σσό			** W	L

/σσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσ		* W	L	*** W
☞ b. σόσσ			*	**
c. σσός			** W	* L
d. σσσό	* W		*** W	L

/σσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
a. όσσσσ		* W	L	**** W
b. σόσσσ		* W	* L	*** W
☞ c. σσόςσ			**	**
d. σσσός	* W		*** W	* L
e. σσσσό	* W		**** W	L

# Two possible refinements

(☹= preferred by generating grammar, losing in acquired grammar)

## \*EXTLAPSE(L) >> \*EXTLAPSE(R)

/σσσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☹ a. όσσσσσσ		*		*****
b. σόσσσσσσ		*	* W	***** L
c. σσόσσσσ		*	** W	**** L
d. σσσόσσ	*! W	*	*** W	*** L
e. σσσσόσσ	*! W	L	**** W	** L
f. σσσσσόσ	*! W	L	***** W	* L
g. σσσσσσό	*! W	L	***** W	L

/σσσσσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	ALIGN(L)	ALIGN(R)
☹ a. όσσσσσσσσ		*		*****
b. σόσσσσσσσσ		*	* W	***** L
c. σσόσσσσσσ		*	** W	**** L
d. σσσόσσσσ	*! W	*	*** W	**** L
e. σσσσόσσ	*! W	*	**** W	*** L
f. σσσσσόσ	*! W	L	***** W	** L
g. σσσσσσόσ	*! W	L	***** W	* L
h. σσσσσσσό	*! W	L	***** W	L

### Midpoint system

2 syl όσ      5 syl σσόσσ  
 3 syl όσσ    6 syl όσσσσ  
 4 syl σσόσσ   7 syl όσσσσσσ

## \*EXTLAPSE(R) >> \*EXTLAPSE(L)

/σσσσσσ/	*EXTLAPSE(R)	*EXTLAPSE(L)	ALIGN(L)	ALIGN(R)
☹ a. όσσσσσσ	*! W		L	***** W
b. σόσσσσσσ	*! W		* L	***** W
c. σσόσσσσ	*! W		** L	**** W
d. σσσόσσ	*! W	*	*** L	*** W
☹ e. σσσσόσσ		*	****	**
f. σσσσσόσ		*	*****! W	* L
g. σσσσσσό		*	*****! W	L

/σσσσσσσσ/	*EXTLAPSE(R)	*EXTLAPSE(L)	ALIGN(L)	ALIGN(R)
☹ a. όσσσσσσσσ	*! W			***** W
b. σόσσσσσσσσ	*! W		* L	***** W
c. σσόσσσσσσ	*! W		** L	**** W
d. σσσόσσσσ	*! W	*	*** L	**** W
e. σσσσόσσ	*! W	*	**** L	*** W
☹ f. σσσσσόσ		*	*****	**
g. σσσσσσόσ		*	*****! W	* L
h. σσσσσσσό		*	*****! W	L

### Antepenultimate stress

2 syl όσ      5 syl σσόσσ  
 3 syl όσσ    6 syl σσσσ  
 4 syl σσόσσ   7 syl σσσσσ

# Ambiguity in short words

- So far: based on  $< 6\sigma$  data, learners exposed to a midpoint system *might* infer antepenultimate stress
  - Hoped-for claim: midpoint system is 'unstable', and learners may learn antepenultimate stress instead
- Problem: ambiguity cuts both ways! Learners exposed to antepenultimate stress might assume that they are learning a midpoint system
  - Actual prediction: variability or changes in both directions
- Where does the antepenultimate bias come from?



# The learning algorithm matters

- RCD doesn't explain antepenultimate bias, because in short words, \*EXTLAPSE(L) and \*EXTLAPSE(R) are 'W-only' constraints  $\Rightarrow$  remain highly ranked
- Stanton's conjecture: human learners actually use a ranking algorithm that doesn't just demote L's, but also promotes W's (Boersma, 1997; Magri, 2012a)

## The learning algorithm matters (*cont.*)

- Why this will help:
  - Short words give lots of evidence for  $\text{ALIGN}(L) \gg \text{ALIGN}(R)$
  - If the learner demotes  $\text{ALIGN}(R)$  and *promotes*  $\text{ALIGN}(L)$ , then  $\text{ALIGN}(L)$  will end up above other markedness constraints
  - Similarly, 4-5 $\sigma$  provide evidence for  $^*\text{EXTLAPSE}(R) \gg \text{ALIGN}(L)$ , causing it to be promoted
  - Consequence:  $^*\text{EXTLAPSE}(L)$  is 'left in the dust' (not promoted until you get 6+ syllable words)

# Learning from short words: promotion and demotion

/σσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	←ALIGN(L)	ALIGN(R)→
☞ a. όσ				*
b. σό			* W	L

/σσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	←ALIGN(L)	ALIGN(R)→
☞ a. όσσ				**
b. σόσ			* W	* L
c. σσό			** W	L

/σσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	←ALIGN(L)	ALIGN(R)→
a. όσσσ		* W	L	*** W
☞ b. σόσσ			*	**
c. σσόσ			** W	* L
d. σσσό	* W		*** W	L

/σσσσσ/	*EXTLAPSE(L)	*EXTLAPSE(R)	←ALIGN(L)	ALIGN(R)→
a. όσσσσ		* W	L	**** W
b. σόσσσ		* W	* L	*** W
☞ c. σσόσσ			**	**
d. σσσόσ	* W		*** W	* L
e. σσσσό	* W		**** W	L

# Learning from short words: promotion and demotion

/σσ/	ALIGN(L)←	*EXTLAPSE(L)	*EXTLAPSE(R)	→ALIGN(R)
☞ a. όσ				*
b. σό	* W			L

/σσσ/	ALIGN(L)←	*EXTLAPSE(L)	*EXTLAPSE(R)	→ALIGN(R)
☞ a. όσσ				**
b. σόσ	* W			* L
c. σσό	** W			L

/σσσσ/	ALIGN(L)←	*EXTLAPSE(L)	*EXTLAPSE(R)	→ALIGN(R)
a. όσσσ	L		* W	*** W
☞ b. σόσσ	*			**
c. σσός	** W			* L
d. σσσό	*** W	* W		L

/σσσσσ/	ALIGN(L)←	*EXTLAPSE(L)	*EXTLAPSE(R)	→ALIGN(R)
a. όσσσσ	L		* W	**** W
b. σόσσσ	* L		* W	*** W
☞ c. σσόςσ	**			**
d. σσσός	*** W	* W		* L
e. σσσσό	**** W	* W		L

# Learning from short words: promotion and demotion

/σσ/	←ALIGN(L)→	*EXTLAPSE(L)	←*EXTLAPSE(R)	←ALIGN(R)→
☞ a. όσ				*
b. σό	* W			L

/σσσ/	←ALIGN(L)→	*EXTLAPSE(L)	←*EXTLAPSE(R)	←ALIGN(R)→
☞ a. όσσ				**
b. σόσ	* W			* L
c. σσό	** W			L

/σσσσ/	←ALIGN(L)→	*EXTLAPSE(L)	←*EXTLAPSE(R)	←ALIGN(R)→
a. όσσσ	L		* W	*** W
☞ b. σόσσ	*			**
c. σσόσ	** W			* L
d. σσσό	*** W	* W		L

/σσσσσ/	←ALIGN(L)→	*EXTLAPSE(L)	←*EXTLAPSE(R)	←ALIGN(R)→
a. όσσσσ	L		* W	**** W
b. σόσσσ	* L		* W	*** W
☞ c. σσόσσ	**			**
d. σσσόσ	*** W	* W		* L
e. σσσσό	**** W	* W		L

# Learning from short words: promotion and demotion

/σσ/	*EXTLAPSE(R)←	←ALIGN(L)→	*EXTLAPSE(L)	←ALIGN(R)→
☞ a. όσ b. σό		* W		* L

/σσσ/	*EXTLAPSE(R)←	←ALIGN(L)→	*EXTLAPSE(L)	←ALIGN(R)→
☞ a. όσσ b. σός c. σσό		* W ** W		** * L L

/σσσσ/	*EXTLAPSE(R)←	←ALIGN(L)→	*EXTLAPSE(L)	←ALIGN(R)→
a. όσσσ ☞ b. σόσσ c. σσόσ d. σσσό	* W	L * ** W *** W	* W	*** W ** * L L

/σσσσσ/	*EXTLAPSE(R)←	←ALIGN(L)→	*EXTLAPSE(L)	←ALIGN(R)→
a. όσσσσ b. σόσσσ ☞ c. σσόσσ d. σσσόσ e. σσσσό	* W * W	L * L ** *** W **** W	* W * W	**** W *** W ** * L L

# \*EXTLAPSE(R) ≫ ALIGN(L) ≫ \*EXTLAPSE(L) ≫ ALIGN(R)

- This ranking works for words of 2–5 syllables
- But it predicts antepenultimate stress for longer words

/σσσσσσ/	*EXTLAPSE(R)	ALIGN(L)	*EXTLAPSE(L)	ALIGN(R)
⊖ a. σσσσσσ	*! W	L	L	***** W
b. σόσσσσ	*! W	* L	L	**** W
c. σσόςσσ	*! W	** L	L	*** W
☞ d. σσσόςσ		***	*	**
e. σσσσόςσ		****! W	*	* L
f. σσσσσόσ		****!* W	*	L

/σσσσσσσσ/	*EXTLAPSE(R)	ALIGN(L)	*EXTLAPSE(L)	ALIGN(R)
⊖ a. σσσσσσσσ	*! W	L	L	***** W
b. σόσσσσσσ	*! W	* L	L	***** W
c. σσόςσσσσ	*! W	** L	L	**** W
d. σσσόςσσσ	*! W	*** L	*	*** W
☞ e. σσσσόςσσ		****	*	**
f. σσσσσόςσ		*****! W	*	* L
g. σσσσσσσόσ		*****!* W	*	L

- Learner trained on both midpoint and antepenultimate patterns learns an antepenultimate grammar
  - ...until long words are encountered, if it's not too late

## Stepping back: the approach, more generally

- Some unattested systems may be possible to capture grammatically, but are difficult to learn
- Goal: theory of grammatical learning that predicts that learners, when exposed to typical input from a 'difficult' pattern, systematically misacquire it as a different, more commonly attested pattern
- Potential to explain not only unattested systems, but also rare systems (which we can't exclude as impossible grammars, anyway)
- Converging evidence: acquisition data, learning in the lab?





# Stress typology



101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

# Modeling typological frequency

Why are some patterns common cross-linguistically, while others are rare or non-existent?

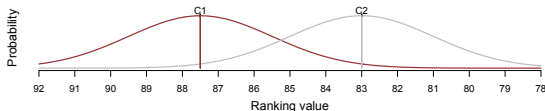
- So far: focus on why certain patterns are unattested
  - Can't be expressed in theory
  - Can be expressed, but difficult to learn
- If learnability is a filter, should affect not only what is attested, but also relative frequencies



# Common vs. rare patterns

## Modeling relative frequency by sampling rankings

- Recall: stochastic ranking (Boersma, 1997; Boersma and Hayes, 2001)
  - Rankings are sampled from distributions at time of evaluation
  - Relative frequency of patterns is determined by ranking distributions



# Predicting typological frequency by sampling rankings

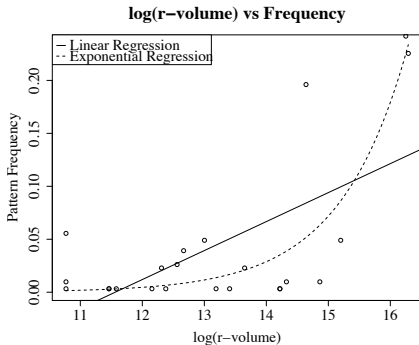
An (overly) simple place to start

- Assume no fixed rankings
- Sample rankings from full factorial typology
- Relative freq of patterns  $\propto$  probability of sampling rankings that generate them
  - If all rankings are equally probable, this equals the *number* of rankings that generate them
  - Bane and Riggle (2008); Riggle (2010): R-VOLUME



# Predicting typological frequency by sampling rankings (*cont.*)

- Positive correlation for freq. of single stress systems: Bane and Riggle (2008, Fig. 6)



# The role of learnability

- ‘Big Bang’ model of typology: distribution generated randomly, stable over time
- In actuality, languages change, in non-random ways
- Reanalysis: patterns are reinterpreted as similar (“confusable”; Bane and Riggle, 2008) patterns
- Reanalysis is directional
  - Certain patterns are difficult to acquire because the evidence that distinguishes them from other “preferred” patterns is rare in the input to learners (Staubs, 2014b; Stanton, 2016)
  - These patterns may be mislearned, becoming rare or even unattested over time

- Staubs, Stanton show that loss through reanalysis is useful for specific gaps in the attested typology (long windows, midpoint systems)
- Broader question: does iterated learning help improve the fit to attested typology, over simple r-volume?
  - Work in progress with Vighnesh Subramaniam (MIT)
  - Preview of result: yes, but discrepancies remain

# QI single fixed stress



101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000



# A fruitful testing grounds

Quantity insensitive fixed stress (Gordon, 2002; Heinz, 2007; Bane and Riggle, 2008)

- Can be modeled with modest number of constraints ( $\approx 12$ )
- Candidates for length  $n$ :  $2^n$  (ignoring primary placement)
- Large enough to be interesting, constrained enough to be tractable
- Databases: Gordon (2002), StressTyp2 (Goedemans and van der Hulst, 2009)



## Focus here: single stress systems

- Granularity: descriptions of secondary stress frequently too coarse to differentiate patterns distinguished only in longer words
- Feasibility: reduced constraint set, much reduced candidate set (linear in length of word)
- Caveat: distinction between single stress vs. secondary stress unmentioned/uncertain may be murky; we return to this issue below



# Reminder of constraints

Gordon (2002); Kager (2012); Stanton (2016)

ALIGN-L, ALIGN-R	Assign one * for each $\sigma$ separating stress from the L/R edge of the word
NONFINALITY	Assign one * for stress on the final $\sigma$
*LAPSE	Assign one * for each sequence of two stressless $\sigma$ 's
*LAPSE-L/R	Assign one * if neither of the initial/final two $\sigma$ 's is stressed
*EXTENDLAPSE	Assign one * for each sequence of three stressless $\sigma$ 's
*EXTLAPSE-L/R	Assign one * if none of the initial/final three $\sigma$ 's is stressed

- Also assume unviolated constraints limiting outputs to single stress
  - CULMINATIVITY: every output has a primary stress
  - \*STRESS: eliminates additional stresses



# Factorial typology

- Factorial typology calculated using OTSoft (Hayes et al., 2013)
- Inputs: words of 2 through 8 syllables
- Output candidates: stress on each syllable
- $8! = 40320$  possible rankings ('grammars'), and  $8!$  logically possible combinations of output candidates ('patterns'/'languages')
- Of these, only 48 patterns are predicted to occur, while 40272 patterns cannot be derived



# Categorical attestation

- Counted fixed primary stress placement in STRESSTYP2
  - Simple positions: final, penultimate, antepenultimate; initial, peninitial, postpeninitial
  - Elaborations: penitital + non-finality
  - Two possible 'midpoint' systems (Içña Tupi, Bhojpuri)
- None of the 40272 underivable patterns are attested
- Overgeneration: 39 patterns are predicted by the constraint set, but unattested
- All 9 attested patterns can be derived, though with very different frequencies

Challenge: does an iterated learning model explain why 39 patterns are unattested, and help predict the relative frequencies of attested patterns?



# Ranking-Based Frequencies



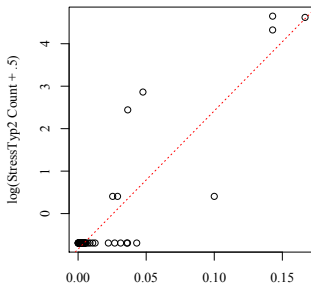
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

## Baseline: r-volume

- Along with Bane and Riggle (2008), Riggle (2010), Staubs (2014b), Stanton (2016) and others, we assume a baseline in which grammars are sampled randomly
- OT: probability of a language  $L$  = its r-volume
$$\frac{\text{Number of rankings that derive } L}{\text{Total number of rankings}}$$
- Calculated r-volume for each of the 48 predicted patterns, using recursive technique proposed by Riggle (2010)

## Comparison to empirical data

- As in Bane and Riggle (2008), compared r-volume with count in STRESSTYP2
- Bane and Riggle suggest that typological frequency may best be modeled by  $\log(\text{r-volume})$
- However, it appears to us that r-volume linearly predicts the log of typological frequency, in line with Poisson models of count data





# Assessment

- Correlation between (log) r-volume and frequencies of 9 attested patterns: Spearman's  $\rho = .612$
- Qualitatively imperfect
  - 39 patterns with non-zero r-volume are empirically unattested
  - Although some of these may be accidental gaps, many are high enough to yield non-zero predicted counts in a database this size (Stanton, 2016)
- Quantitatively imperfect
  - Correlation artificially inflated by large number of unattested patterns
  - Relative frequencies of attested patterns are not well predicted



# Hypothesis: iterated learning could improve both aspects of the model

- Difficult to learn patterns are reanalyzed and become rarer or even unattested over time
- Probability mass reassigned, changing relative predicted relative frequency of attested patterns



# Learning-Based Frequencies



101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

In order to test this hypothesis, we submitted each of the 48 derivable stress patterns to an iterated learning model, to see how probabilities are reassigned over time



# The learning model

- Gradual Learning Algorithm (Boersma, 1997), as modified by Magri (2012b) to guarantee convergence
- Error-driven learning
  - Model receives input/output pairs: /σσσσ/, [σσόσ]
  - Model uses current grammar to derive predicted output
  - If incorrect, adjust ranking values to favor trained output
- Parameters
  - All markedness constraints start out equally ranked (100)
  - Plasticity: starts at 1, gradually decreases to .001 (learning slows with age; step=.004)
  - Learning trials: 2000



# Training

- Trained the learning model on each of the 48 stress patterns
- Inputs: words of 2–8 syllables
- Shorter words presented more often than long words, according to mean frequencies of word lengths given by Stanton (2016)

2 $\sigma$	32%	6 $\sigma$	1%
3 $\sigma$	18%	7 $\sigma$	.1%
4 $\sigma$	9%	8 $\sigma$	.1%
5 $\sigma$	3%		

- Each pattern run 1000 times, yielding a probability distribution over learning outcomes



# Iterated learning

- Starting point: each pattern assigned probability according to baseline distribution (r-volume)
- Imperfect learning changes the probability distribution—e.g., when trained on Pattern 19 (antepenult.  $\leq 5\sigma$ ; else peninit.), the model learned:

Pattern 19	(antepenult $\leq 5\sigma$ ; else peninit)	30.0%
Pattern 20	(antepenult. $\leq 5\sigma$ ; else postpeninit)	34.1%
Pattern 21	(antepenultimate)	25.4%
Pattern 18	(antepenult. $\leq 5\sigma$ ; else init)	8.9%
Pattern 23	(antepenult $\leq 4\sigma$ ; else peninit)	.9%
Patterns 24, 7	(initial, etc.)	<1%

- Generations
  - Baseline (r-volume) distribution reassigned according to learned distributions
  - Iterated for 100,000 generations

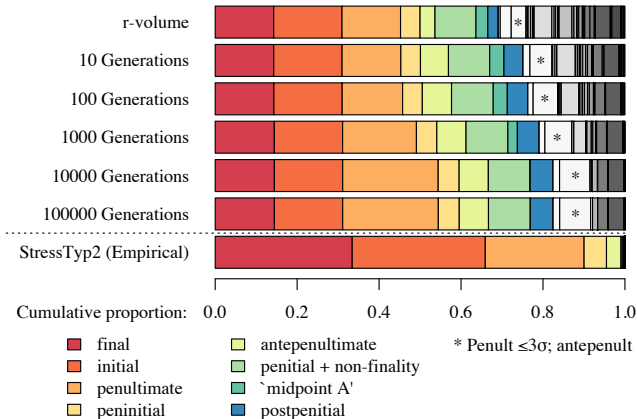


# Results



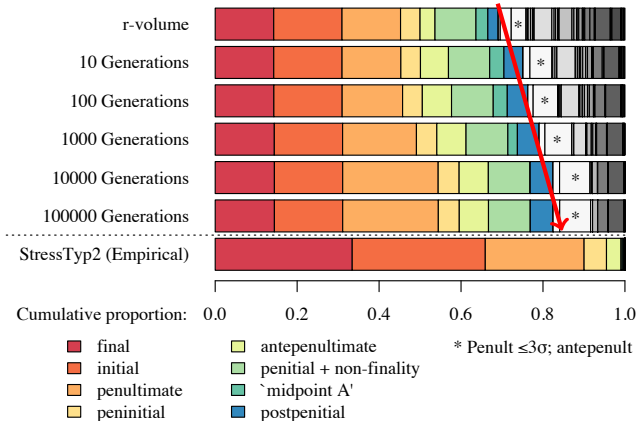


# Promising: reduction of unattested patterns



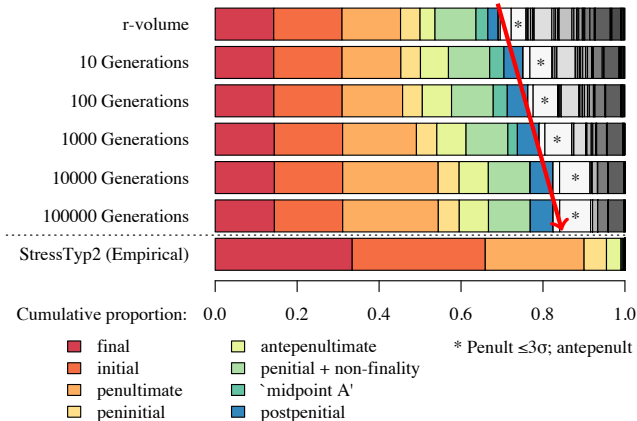
- Patterns attested in STRESSTYP2 (colored bars) increase in frequency over time

# Promising: reduction of unattested patterns



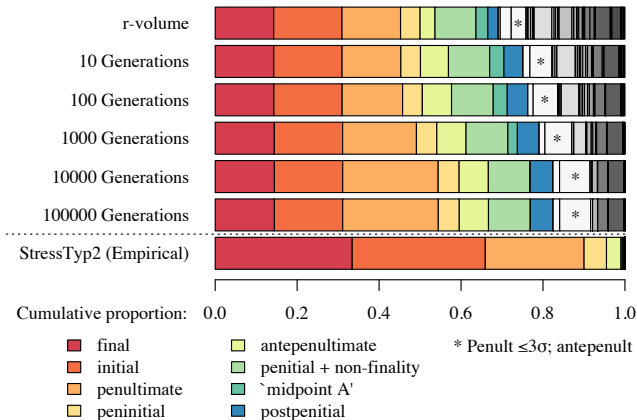
- Patterns attested in STRESSTyp2 (colored bars) increase in frequency over time

# Promising: reduction of unattested patterns



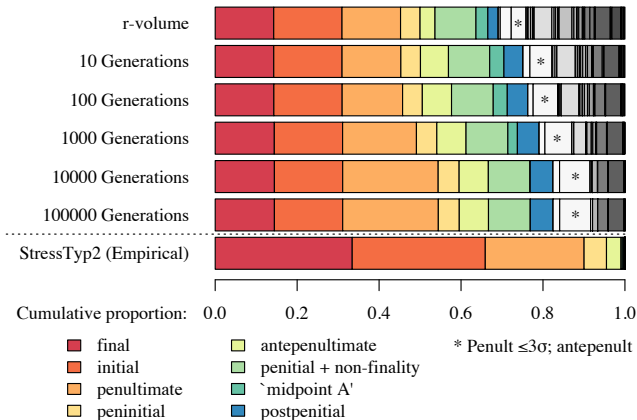
- Unattested patterns (gray bars) decrease in frequency over time

# Promising: reduction of unattested patterns



- Some rare or controversially attested patterns also successfully nearly eliminated ('midpoint A')

# Promising: reduction of unattested patterns



- Consistent with hypothesis that unattested patterns may be eliminated by learning (Blevins, 2004; Staubs, 2014b; Stanton, 2016)

## Discrepancy 1: stubborn unattested patterns

- Although unattested patterns gradually decrease in relative frequency, some remain or even increase
- Example: '\*' penult  $\leq 3\sigma$ , antepenult longer (a 'midpoint system')

\*LAPSE, \*EXTLAPSE, \*EXTLAPSE-R, NONFIN

>>

ALIGN-L, \*LAPSE-L, \*EXTLAPSE-L

>>

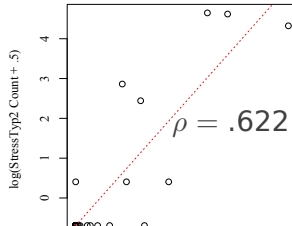
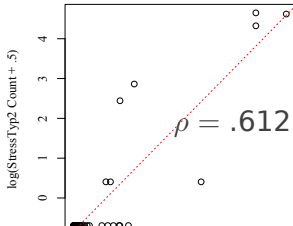
ALIGN-R, \*LAPSE-R

- This pattern starts out with non-negligible probability (r-volume), and increases as other patterns are reanalyzed as it (does not rely heavily on long words)



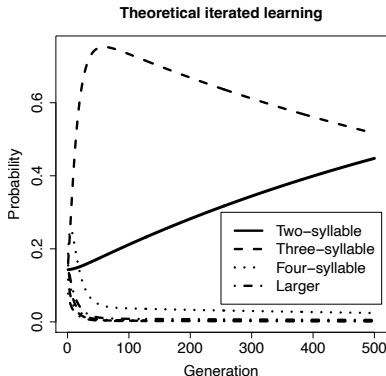
## Discrepancy 2: unexpected beneficiaries

- Among attested (colored) patterns, predicted final distribution is *less* like attested (bottom row) than in the baseline (r-volume)
- Example: peninitial stress with non-finality ('σσ, σ'σσ)
  - Scarcely attested (Southern Paiute), but easily learned
- Post-peninitial stress is also a popular reanalysis for midpoint systems
- Result: no clear improvement in predictions for relative frequency of attested patterns



## A similar effect from Staubs (2014a)

Fig 2.8 (p. 96)



- As longer windows are reanalyzed, frequency of antepenultimate stress increases dramatically



## Local summary

- An attempt to scale up to a (slightly) more system-wide test of an idea that is intuitive and has proven useful in specific cases
  - Iterated grammatical learning winnows some patterns from the attested typology
  - Reanalysis shapes the relative frequency of attested patterns
- Modest support for winnowing, though not the whole story
  - Some unattested patterns remain; explanation for gap must lie elsewhere
- No clear support for learnability-based redistribution
  - In fact, r-volume remains marginally better predictor of typological frequency



What does a modeler do in this situation?



# Wrong constraints?

- Hyde (2008, 2015): redefine ALIGN
- Kager (2012): reject \*LAPSE because it can generate midpoint systems
  - Stanton (2016) addresses some but not all of these
- Switch to foot-based constraints?
  - Could eradicate 'Penult  $\leq 3\sigma$  else antepenult' problem
  - May incorrectly rule out potential midpoint-type systems (Içña Tupi, Bhojpuri)
  - Role of learnability far more difficult to explore: Hidden Structure (Tesar and Smolensky, 2000; Jarosz, 2013; Boersma and Pater, 2016)



## R-volume as a prior?

- R-volume is not perfect, but striking that it does as well as it does. What creates this distribution?
- Bayesian inference:  $P(H|D) \propto P(D|H) * P(H)$ 
  - Suppose 'H' is 'grammar generating pattern  $n$ ' (aggregate over equivalent rankings)
  - Learners may retain H, even when likelihood ( $P(D|H)$ ) is not especially high
  - Acoustic ambiguity of stress makes it especially prone to strong priors?
- Yang, Albright and Feldman (2022) Assessing the learnability of process interactions using grammatical spaces. (*Cogsci* paper)



## A couple lessons

- The importance of scaling up
  - Prosthetic thought: calculating typology and r-volume, multiple runs of iterated learning
  - Results are not always what one expects
- Baselines
  - Results here aren't a particularly accurate model of typology, but they help point to some areas for future improvement
  - Empirical: accuracy of relative frequencies in STRESSTYP?
  - Attention to mismatches may point to changes in the constraints or learning model



## References

- BANE, MAX and JASON RIGGLE. 2008. Three correlates of the typological frequency of quantity-insensitive stress systems. In Proceedings of the Tenth Meeting of ACL Special Interest Group on Computational Morphology and Phonology, 29-38. Association for Computational Linguistics.
- BLEVINS, JULETTE. 2004. Evolutionary Phonology: The Emergence of Sound Patterns. Cambridge University Press.
- BOERSMA, PAUL. 1997. How we learn variation, optionality, and probability. Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21, 43-58.  
<http://fon.hum.uva.nl/paul/>.

## References (*cont.*)

- BOERSMA, PAUL and BRUCE HAYES. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1), 45–86.
- BOERSMA, PAUL and JOE PATER. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic Serialism and Harmonic Grammar*, ed. by John McCarthy and Joe Pater, 389–434. Sheffield: Equinox.
- GOEDEMANS, ROB and HARRY VAN DER HULST. 2009. Stresstyp: A database for word accentual patterns in the world's languages. *The use of databases in cross-linguistics research* 235–282.
- GORDON, MATTHEW. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20(3), 491–552.

## References (*cont.*)

- HAYES, BRUCE; BRUCE TESAR; and KIE ZURAW. 2013. OTSoft 2.5 software package.  
<http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- HEINZ, JEFFREY. 2007. Inductive learning of phonotactic patterns. Ph.D. thesis, UCLA.
- HYDE, BRETT. 2008. Alignment continued: distance-sensitivity, order-sensitivity, and the midpoint pathology. ROA 998.
- . 2015. The midpoint pathology: What it is and what it isn't. ROA 1231.
- JAROSZ, GAJA. 2013. Learning with hidden structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30, 27–71.



## References (*cont.*)

- KAGER, RENÉ. 2012. Stress in windows: Language typology and factorial typology. *Lingua* 122, 1454–1493.
- MAGRI, GIORGIO. 2012a. Convergence of error-driven ranking algorithms. *Phonology* 29(2), 213–269.
- . 2012b. Convergence of error-driven ranking algorithms. *Phonology* 29(2), 213–269.
- PRINCE, ALAN and PAUL SMOLENSKY. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.
- RIGGLE, JASON. 2010. Sampling rankings. ROA-1075 .
- STANTON, JULIET. 2016. Learnability shapes typology: the case of the midpoint pathology. *Language* 92(4), 753–791.

## References (*cont.*)

- STAUBS, ROBERT D. 2014a. Computational modeling of learning biases in stress typology. Ph.D. thesis, University of Massachusetts, Amherst.
- STAUBS, ROBERT D. 2014b. Computational modeling of learning biases in stress typology .
- TESAR, BRUCE and PAUL SMOLENSKY. 2000. Learnability in Optimality Theory. Cambridge, MA: MIT Press.