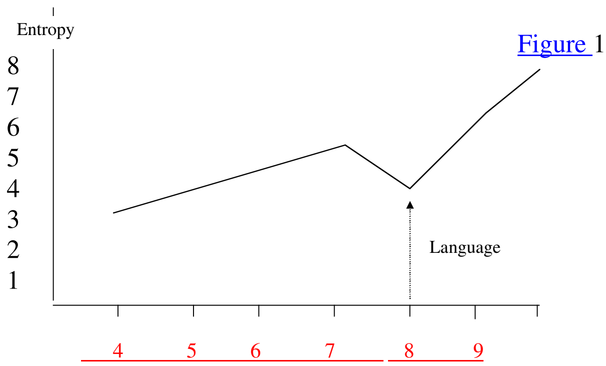

Entropy profile as an indicator of character bit-length

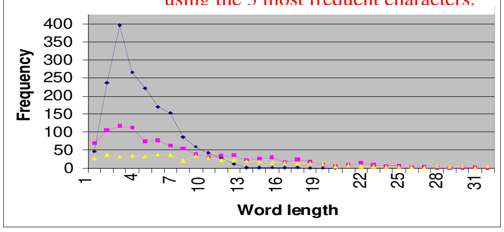

Figure 2: Candidate word-length distributions using the 3 most frequent characters.

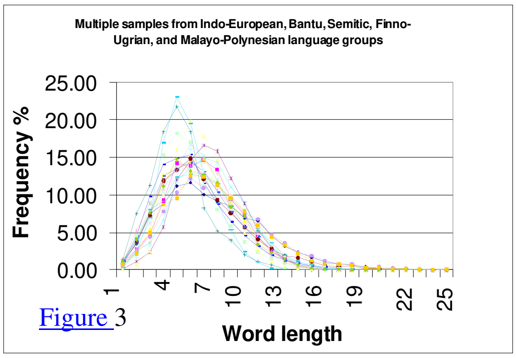

Figure 3

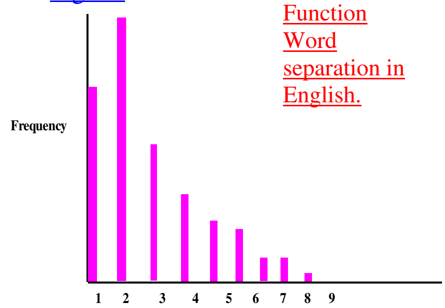

Figure 4

Number of words between candidate functional words

Correlation profile for word pair

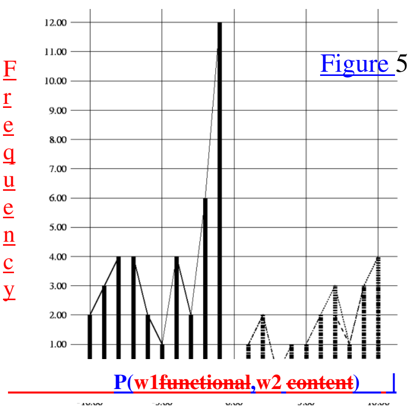

Figure 5

Figure 6: VB-tag profile

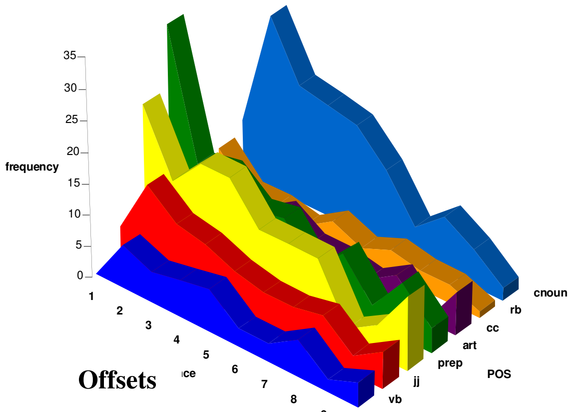

Figure 7	Cnoun	Jj	Rb	Prep	Cc	Vb	Art
Cnoun	β, λ_3	δ*	λ_2	β*	β*, λ_6	δ, λ_2	δ, λ_2
Jj	β*	β	$\delta, \lambda_{5,9}$	λ_2	$\lambda_{2,4}$	δ	δ, λ_3
Rb	Z, λ_5	λ_7	β	β*	δ, λ_9	β	λ_2
Prep	δ*, λ_2	λ_2	δ*, λ_7	δ, λ_3	λ_3	Z*,λ_9	β
Cc	δ*, $\lambda_{3:4}$	β	β, Z_6	λ_4	Z	λ_5	β*
Vb	λ_2	λ_2	β	β*	δ, Z_9	Z	β*
Art	β*	β*	$\delta, \lambda_{3,8}$	Z, λ_2	Z*	Z	Z, λ_4

Key: Z = Zero bigram - or at offset specified - occurrences. δ = Very weak bonding - near zero - at bigram occurrences. β = Strong bonding at bigram co-occurrences. * = Indicates opposing cohesive trend when P.O.S. reversed. λ_n = High peak beyond bigram at offset distance of n . Φ = Flat distribution across offsets - bigram bonding evident.