Background
• Alzheimer’s Disease (AD) and other types of dementia are associated with changes in spoken language.
• Language Evaluation is time consuming and in most cases subjective.

Aims of the study
• To adopt a computational approach based on machine learning (ML) to analyze language samples from native speakers of English and Greek in order to automatically detect early indicators of AD.
• To identify AD-induced language characteristics that are either cross-linguistic or language-specific.

Materials and Methods
Speech samples were taken from three sources: Two from archived (English) language resources and one (Greek) collected for this project. Participants were shown the “cookie theft” picture and were asked to describe what they could see happening.

Analytical Approach
Feature Extraction: Bag of Words assumption (BoW), Part of Speech (PoS) tags, Lexical Variation (LV) and Syntactic Complexity (SC) measures.

Feature Selection: (1) Common top ranked Information Gain (IG) words across the three data sets; (2) Commonly distinctive PoS, LV, SC between AD & NC groups (p-value < 0.05) across data sets.

Classification of spoken samples

Data sets
Class: Samples (Aug. MMSE)

<table>
<thead>
<tr>
<th>DEMENTIA (USE)</th>
<th>OPTIMA (UZ)</th>
<th>GREEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD: 39 (19.0)</td>
<td>AD: 18 (21.3)</td>
<td>AD: 17 (20.0)</td>
</tr>
<tr>
<td>NC: 246 (27.5)</td>
<td>NC: 248 (27.0)</td>
<td>NC: 14 (28.6)</td>
</tr>
</tbody>
</table>

Cross Lingusitic Analysis Results
(1) Correlations/overlap of common top-100 ranked (IG) words

<table>
<thead>
<tr>
<th></th>
<th>DEMENTIA-GREEK</th>
<th>DEMENTIA-OPTIMA</th>
<th>OPTIMA-GREEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s Correlation</td>
<td>0.3</td>
<td>0.79</td>
<td>0.52</td>
</tr>
<tr>
<td>Spearman’s Rank Correlation</td>
<td>0.26</td>
<td>0.62</td>
<td>0.39</td>
</tr>
</tbody>
</table>

% Common words in top-100: 0.49, 0.36, 0.41

(2) Commonly distinctive (t-Test, p<0.05) PoS, LV, SC features between AD & NC groups across data sets

<table>
<thead>
<tr>
<th>Crosslingual (i.e. all samples)</th>
<th>US English &amp; Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CCW</strong> (AD&gt;NC)</td>
<td>Adverb Freq. (AD&gt;NC)</td>
</tr>
<tr>
<td><strong>CWR</strong> (AD&gt;NC)</td>
<td>Nouns/Tokens (NC&gt;AD)</td>
</tr>
<tr>
<td>Nouns Freq. (AD&gt;NC)</td>
<td>Pronouns/Nouns (AD&gt;NC)</td>
</tr>
<tr>
<td>Mean Length Sentence (NC&gt;AD)</td>
<td></td>
</tr>
</tbody>
</table>

*Closed Class Words Count (conjunctions, determiners, prepositions, pronouns)
**Closed Class/Open Class Words Ratio (Open Class: nouns, adjectives, adverbs)

Conclusions
• Discriminative power of LV, SC and PoS features verified across languages.
• State-of-the-art accuracy of deployed system.
• Development of CogAware prototype.

Future Work
• Fusion of crosslinguistic features with best model (complementarity of errors)
• Preliminary findings indicate enhanced performance

Additional Work
- CogAware App
- Automatic transcription of spoken sample
- Research prototype for automatic language-based assessment of AD risk factor
- Android application and Google Drive Add-on
- Quick and accurate screening of patients for AD
- Pilot deployed in day-care centers (GR)

References
6. The Natural History of Alzheimer’s Disease: Description of Study Cohort and Accuracy of Diagnosis. Archives of Neurology, 51(6), 858-859.