# Radical Allomorphy: Phonological Surface Forms without Phonology

**Salam Khalifa,[1,2] Nizar Habash,[2] Owen Rambow[1]**
Institute for Advanced Computational Science,
[1]Stony Brook University
Computational Approaches to Modeling Language (CAMeL) Lab,
[2]New York University Abu Dhabi
{first.last}@stonybrook.edu, {first.last}@nyu.edu

## Abstract

Recent computational work typically frames morphophonology as generating surface forms (SFs) from abstract underlying representations (URs) by applying phonological rules or constraints. This *generative* stance presupposes that every morpheme has a well-defined UR from which all allomorphs can be derived, a theory-laden assumption that is expensive to annotate, especially in low-resource settings. We adopt an alternative view. Allomorphs and their phonological variants are treated as the basic, observed lexicon, not as outputs of abstract URs. The modeling task, therefore, shifts from deriving SFs to selecting the correct SF, given a meaning and a phonological context. This discriminative formulation eliminates the need to posit or label URs, allowing the model to exploit surface evidence directly.

## 1 Introduction

Computational morphophonology has long modeled the problem of deriving spoken forms, also known linguistically as surface forms (SFs), from underlying representations (URs) using some form of transformations. Whether within rule-based frameworks, constraint-based theories, or neural sequence-to-sequence approaches, this *generative* tradition hinges on the idea that URs exist as abstract, latent forms from which observable variants, or allomorphs, can be derived. Yet despite the large amount of work invested in this problem, the status of the UR remains conceptually and practically fraught: What constitutes a valid UR? How are its properties discovered, and how does one distinguish it from a mere analytical convenience? Is it suitable when observable data is scarce, where there aren't enough examples to hypothesize a UR?

In this paper, we explore a radically different approach. Rather than positing URs and seeking generative mechanisms to produce SFs, we propose a model in which the full set of surface variants, the combination of the allomorphs and their phonological surface variants (which we will call *allomorphones*), is taken as primary. The task then becomes one of enumeration and selection: given a context, choose the appropriate SF from a prespecified inventory. This reframing transforms the problem from a generative to a discriminative one, with significant implications for both linguistic theory and computational modeling. Figure 1 illustrates how multiple allomorphones of the +Plural morpheme for English coexist in the surface set, with one selected per context.

| Root | +Plural | | | SF |
|---|---|---|---|---|
| | UR | SS | SF | |
| dog | +s | {+s, +z, +ɪz} | +z | [dɔgz] |
| cat | +s | {+s, +z, +ɪz} | +s | [kæts] |
| fox | +s | {+s, +z, +ɪz} | +ɪz | [fɑksɪz] |
| ox | +en | {+en} | +en | [ʔksən] |

Figure 1: Examples of English plural morphology illustrating the distinction between underlying representations (UR), the surface set of possible realizations (SS), and the selected surface form (SF).

By grounding analyses in observable surface data rather than abstract representations, our approach emphasizes empirical adequacy and learnability. It aligns with recent computational trends favoring surface-level, context-sensitive modeling. The approach is particularly exciting in low-resource settings for which a limited amount of data is available, as creating a UR requires theoretical decisions about the phonological system of the language or language variant, while our approach only requires morphological segmentation and part-of-speech annotation. We demonstrate its utility through a case study in Egyptian Arabic for its complex morphophonology. We compare our approach with existing models analogous to generative approaches. We also show that our model excels in low-resource settings compared to others.



*Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 20435–20441
November 4-9, 2025 

## 2 Background and Related Work

### 2.1 Linguistic Theory

In theoretical linguistics, there is extensive literature discussing different theories of the mental representation assigned to morphemes. On one end, Generative Phonology (**GP**; Kenstowicz and Kisseberth, 1979), assumes one abstract UR per morpheme and derives all surface forms through language-specific phonological rules or constraints, this derivation is a transformation of the UR into the SF. For example, within the rule-based framework, each rule accounts for a single transformation, therefore, there is often a specific order that governs the application of the rules to a UR to generate the correct SF. Similarly, in a constraint-based framework, multiple constraints are lined up according to a specific rank to pick the correct hypothesized SF from a given UR. These ordered rules and ranked constraints are often language- or dialect-specific and require extensive linguistic scholarship. On the other end, researchers argue against having a single abstract representation and instead lexically store every context-conditioned alternant and allow phonological mechanisms to choose the appropriate form, leaving little work for phonological rules (Harris, 1942; Becker and Gouskova, 2016). This framework is referred to as the Morpheme Alternate Theory (MAT). While this approach does not posit an abstract representation, it still requires language-specific mechanisms. More extensive discussion on the implications of those views is found in Hwangbo (2018, 2023).

In this work, we don't necessarily subscribe to one specific theory over the other. Instead, we empirically investigate the implications of each approach for the task of recovering SFs from a given representation regardless of the theoretical grounding of such a representation.

### 2.2 Computational Modeling

The task of deriving a spoken form of a word in NLP has been approached from a generative-like perspective, where an initial form goes through transformations to arrive at the spoken form. Grapheme-to-phoneme (G2P) tasks consider such form to be the orthographic representation; the latest SOTA systems are mostly encoder-decoder based models (Ashby et al., 2021; McCarthy et al., 2023) where the transformation mechanisms are often uninterpretable and opaque. Other works that model morphophonology either assume a UR or discover one based on some constraints (Antworth, 1991; Habash and Rambow, 2006; Pater et al., 2012; Cotterell et al., 2015; Belth, 2023; Khalifa et al., 2023).

This work is inspired by the long-standing literature in computational phonology, which aims to model existing grammars and constraints efficiently. One notable work is the One-Level Phonology by Bird and Ellison (1994). In their work, they compile well-formedness constraints as "state-labeled" automata, which are then combined with logical operations regardless of ordering. Most importantly, they do not rely on the concept of UR and instead operate directly on surface forms. While this is similar to our work in terms of using surface forms only, our work extracts selection constraints (akin to well-formedness) from the data itself with very few priors, unlike One-Level Phonology, where these constraints are based on linguist-provided descriptions. Moreover, our approach is trained and tested on a transcribed corpus of naturally occurring speech, which is a more realistic application.

To the best of our knowledge, we have not encountered efforts approaching this task as an enumeration and selection problem. We believe this is an important path to explore to facilitate spoken form derivation for low-resource and understudied language varieties. We explore our approach through studying Egyptian Arabic morphophonology, since it is inherently a morphologically complex language variety and well studied by phonologists. For this reason, we consider PARLA (Khalifa et al., 2023) to be the computational model analogous to GP that we compare to. We use the comprehensive dataset described in their work. In short, PARLA is a rule-inducing algorithm that creates and generalizes productive transformation rules from a set of UR-SF pairs. Those rules represent the morphophonological grammar that generalizes to new unseen URs. PARLA operates within the rule-based phonology framework but is not an absolute representative of it.

## 3 Our Approach: Radical Allomorphy

In this approach, we shift the focus from the direct relationship between the actual form of the underlying representation to what it represents: a set of realized surface forms. We explore an approach analogous to MAT, where we don't assume a phonological UR and instead represent the dif-

| MTAG | CONJ:f | DET | NOUN | FS | # | NOUN | FS.Cnstr | POSS:1P |
|---|---|---|---|---|---|---|---|---|
| HMORPH | fa- | il- | sAH | =a | # | sAH | =it | =hum |
| ALM | {**fa-**, f-} | {il-, is-, it-, ir-, ⋯ l-, **s-**, t-, r-, , ⋯} | {**sAH**, saH} | =a | # # | {sAH, **saH**} | {**=it**, =t} | {**=hum**, =uhum} |
| SF | fa- | s- | sAH | =a | # | saH | =it | =hum |

Table 1: Example illustrating the taxonomy of our radical allomorphy approach: Morph tag (MTAG), Head Morph (HMORPH), Allomorphones (ALM), and Surface Form (SF). The example corresponds to the Egyptian Arabic sentence فالساحة ساحتهم [fassa:Ha saHithum] 'so the yard is their yard'. '-', and '=' indicate prefix, and suffix boundaries, respectively.

ferent realizations of morphs as a set of possible SFs. By this definition, the task becomes one of enumeration and selection rather than an explicit mapping between a form and its realization.

We define three key notions:

- **Morph Tag (MTAG)** is the morphosyntactic identity of a morpheme, i.e., a set of feature-value pairs along with an underlying part-of-speech tag (e.g., verb, 3rd person singular masculine perfective).
- **Head morph (HMORPH)** is a single morpheme that represents the set of its corresponding surface form realizations without necessarily being related to them by phonological processes.
- The **Allomorphones (ALM)** form the set of surface form realizations of a certain morph.

This approach, naturally, assumes the data to be a parallel set of segmented SFs with their corresponding MTAGs and possibly HMORPHs. This allows the extraction and clustering of the ALM sets and linking them to the different levels of representations.

We illustrate this terminology in Table 1. In practical terms, the sets of ALMs are formed by leveraging the alignment between the segmented SFs in the data and their corresponding MTAGs and/or HMORPHs. Multiple morphophonological phenomena are shown in the example: elision of /i/ after /a/, determiner /il-/ assimilation, and unstressed long vowel shortening. In our approach, these rules are not explicitly described or modeled in order to select the desired SF. Instead, we show that generic selection heuristics work well.

### 3.1 Representations

We empirically explore the effectiveness of our approach through evaluating different ways of representing the morphs. This means that the starting point of the clustering of the set of ALM is either the MTAG level or the HMORPH level.

**MTAG > ALM** In this setup, the MTAG is the morphological tag associated with each morph, which we can think of as the meaning of the morph. For clitics that have identical tags (primarily particles, conjunctions, and prepositions), the tag is additionally augmented by its lexical form, in this case the HMORPH. Stems on the other hand, will have their MTAG tag augmented by the vowelized templatic shape of the HMORPH.

**HMORPH > ALM** A simpler setup where one form of the morphs is the starting point. For affixes and clitics, we chose the HMORPH to be what was previously treated as the UR by PARLA, this allows for direct comparison. However, nothing hinges on this specific choice, and other choices are also possible. For stems, the HMORPH is the vowelized template of the selected HMORPH. Crucially, we do not distinguish between morphs that share the same meaning but have the same HMORPH.

**MTAG-HMORPH > ALM** This is a variant of MTAG > ALM in which the HMORPH becomes part of the MTAG to see the effect of having a middle-ground starting point.

For all approaches, the stem abstraction is also applied to the corresponding ALM sets. This is to facilitate generalization by filling possible gaps across sets of ALM that share similar templatic HMORPH. This intuition is based on the templatic properties of stems in Arabic. In the example in Fig 1, the HMORPH of the noun is CAC and the ALM set is {CAC, CaC}, so if another stem with the same template such as the stem of ساعة /sa:ʕa/ 'hour' has one ALM realization in the data, say CAC, the missing CaC will be recovered since both entries will map into the same HMORPH space.

### 3.2 Core Approach

We now describe our *enumeration* and *selection* approach. The input is a sequence of morphs represented in one of the three ways just described.

**Enumeration** Given an input, at any level, a set of candidate SFs is generated through combining the members of the corresponding ALM set for each morph, thus, *enumerating* every possible combination. For example, for the second word in Figure 1 ساحتهم [saHithum], if the input was in the HMORPH level: `sAH=it=hum` the possible SF candidates will be {`sah=it=hum`, `sAH=t=uhum`, `sAH=t=hum`, `saH=t=hum`, ... etc}. We built the generator [1] using the PyFoma toolkit (Hulden et al., 2024).[2] The FSTs were built offline based TRAIN.

**Selection** We employ four different selection criteria inferred from TRAIN. **a)** Well-formedness of the candidate syllabic form based on the syllabic structure grammar in the training data. This eliminates ill-formed candidates such as `sAH=t=hum` and `saH=t=hum` since they exhibit a `CCC` cluster which is not found in Egyptian Arabic, **b)** The likelihood of the SF syllabic structure given the input, **c)** A morph-based uni-gram model of the ALMs, i.e., a candidate SF with more frequent ALMs will be favored over one with rarer ALMs, **d)** A character based tri-gram model over the full SFs, where we only consider tri-grams with boundaries in them.

Both **a)** and **b)** take into account the whole word, while **c)** focuses on morphs, and **d)** on possible phonological changes around the morpheme boundaries. Moreover, **a)** is discriminative, while **b)**, **c)**, and **d)** are likelihood models that produce scores that are then summed without weights to get a final score. These criteria are based on observations about the interaction between morphophonology and syllabic structure constraints in Egyptian Arabic, and Arabic in general (Broselow, 1976, 2017).

It is worth noting that these selection criteria are not to be confused with the rules in rule-based GP, which are linguistically-specific *rewrite* transformations applied when the application context matches. These criteria are generic and are automatically set based on the data.

## 4 Evaluation and Discussion

### 4.1 Data

Following Khalifa et al. (2023) for purposes of comparison, we use the same dataset they described in their work, which consists of pairs of UR-SF along with their fine-grained morphological tag, which was never used in their work. We also follow their splits and their reporting performance on out-of-vocabulary (OOV) entries only to truly evaluate generalization capabilities. The splits are as follows TRAIN (12,658 types), OOV-DEV (2,190 types), and OOV-EVAL (2,271 types).

The SFs in the original dataset were not morphologically segmented. The segmentation for SFs in TRAIN were achieved by projecting the morpheme boundaries in the URs onto their corresponding SFs through character based alignment. In cases were multiple projections are plausible, we consider all of them and add all possible allomorphones to the ALM set.

### 4.2 Baselines

In addition to our systems mentioned above, we evaluate on three baselines: **a) HMORPH > HMORPH**: Where SF = HMORPH, **b) NEURAL**: A SOTA neural character-based transformer (Wu et al., 2021) that we train to generate the SF from a given HMORPH, and finally **c) PARLA**, for which we compare to the authors' latest reported results in (Khalifa et al., 2025).

Results of our evaluation are shown in Table 2. Apart from the NEURAL baseline, the best performing system is our HMORPH > ALM configuration followed by PARLA. Between our three proposed setups, the configuration with no explicit morphosyntactic identity (HMORPH > ALM) is the best performer. The purely morphological configuration (MTAG > ALM) performs worst, but just slightly worse than PARLA. The MTAG-HMORPH > ALM improves slightly on the purely moprhological system. We conducted further analysis which revealed that using the MTAG information splits and collapses the sets of ALM that either share the form (space is split) or share the meaning (space is collapsed). This dichotomy hindered the generalizability in the affix and clitic space.

For our systems, we also report the hard upper bound of the performance, the oracle, which is simply the presence of the correct answer among the generated candidates. This evaluation highlights the following: 1) in all our setups the oracle is extremely competitive with the NEURAL baseline, 2) the effect of the MTAG is mostly on the selection criteria rather than generation which suggests the importance of the morph based uni-gram model which fully depends on the distribution of the members within an ALM set, and as such splitting or merging such sets will affect such model. These findings clearly suggest that having a stronger se-

[1] "Generation" here means "production", *unrelated* to GP.
[2] We use release v1.0.7

| System | OOV-DEV | OOV-EVAL |
|---|---|---|
| HMORPH > HMORPH | 37.1 | 36.0 |
| PARLA'25 | 81.6 | 80.8 |
| NEURAL | 92.9 | 91.4 |
| MTAG > ALM | 78.4 (92.5) | 78.6 (91.5) |
| HMORPH > ALM | **85.2** (93.0) | **84.5** (91.8) |
| MTAG-HMORPH > ALM | 79.6 (92.3) | 80.0 (91.3) |

Table 2: Accuracy of predicting the correct SF. For our systems, we report the *oracle* results (in parentheses) where the correct answer has been generated but not necessarily selected.

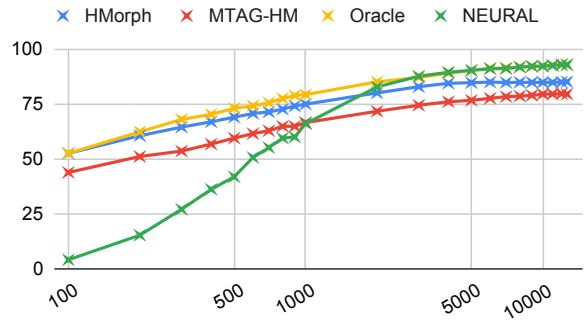


Figure 2: Accuracy on OOV-DEV for all systems across different sets of training sizes.

lection approach is key to a better performance and subsequently a better and explainable model of morphophonology.

To further evaluate the generalizability of our models compared to the absolute best performer, NEURAL, we conducted a learning curve experiment where we split the TRAIN into low- and mid-resource sizes to simulate real-life scenarios of dialects with impoverished resources. The results are shown in Figure 2. The x-axis (in log scale) represents the different training sizes: from 100-1,000 with increments of 100 to represent the extremely low-resource, and then from 1,000 to full TRAIN to represent mid- to high-resource. The y-axis is the accuracy in percentages. Our systems outperform NEURAL in the very low-resource setting. By around 2,000 samples, NEURAL picks up. It is also worth noting our best system outperforms PARLA with only 3,000 samples which makes it a more suitable system for low- to mid-resource scenarios. Detailed percentages are in Table 3 in the appendix.

## 5 Conclusion and Outlook

In this work, we empirically explored a novel approach to the task of modeling morphophonology. Instead of positing underlying representations or learning mappings from a rigid abstract representation to generate a spoken form, we proposed a rather discriminative approach to select the correct surface form from an enumerated set of candidate surface forms. We find that the most radical allomorphy approach that employs morphological tags as identities of the morphs is more restrictive than a purely form-based allomorphy approach.

We also show that our best setup outperforms a rule-inducing SOTA using around 25% of the data only. Additionally, our approach outperforms the SOTA character-based transformer in the same task in the low-resource scenario.

From a theoretical perspective, adopting a Generative Phonology or a Morpheme Alternate Theory based approach will have empirical implications on the quality of the resulting spoken forms and the representation of the acquired grammar, as our work shows. This sets the path for further exploration on modeling such theories computationally and what it means for practical applications. From a practical NLP point of view, the main takeway is that our approach generates spoken forms of words efficiently at low-resource settings and acquires an interpretable representation.

We plan to explore additional starting points and selection criteria in addition to adding a learning component that extrapolates allomorphones in a way inspired by computational models of language acquisition. We also plan to evaluate this approach on different dialects of Arabic as standalone varieties or cross-dialectal evaluation.

## Limitations

In this work, we acknowledge the following limitations:

- **Surface Form morphological segmentation**: We are aware that the task of morphological segmentation in general is an open research question. As such, we plan to empirically investigate in the future how different alignment and segmentation techniques could affect the performance of our approach.
- **Choice of HMORPH**: To facilitate comparison with the baseline system PARLA, we opted to make the choice of our HMORPH to be aligned with their choice of UR. However, this is another open question to which the answer would be to experiment with different choices of forms that act both as the HMORPH in our case and the UR in PARLA.

## Acknowledgments

We thank Michael Becker, Ellen Broselow, Jordan Kodner, and Robert Hoberman for their helpful discussion and efforts. We thank the anonymous reviewers for their valuable feedback. This research was carried out on the High Performance Computing resources at New York University Abu Dhabi. Rambow gratefully acknowledges support from the Institute for Advanced Computational Science at Stony Brook University.

## References

Evan L Antworth. 1991. Introduction to two-level phonology. *Notes on Linguistics*, 53:4–18.

Lucas F. E. Ashby, Travis M. Bartley, Simon Clematide, Luca Del Signore, Cameron Gibson, Kyle Gorman, Yeonju Lee-Sikka, Peter Makarov, Aidan Malanoski, Sean Miller, Omar Ortiz, Reuben Raff, Arundhati Sengupta, Bora Seo, Yulia Spektor, and Winnie Yan. 2021. Results of the second sigmorphon shared task on multilingual grapheme-to-phoneme conversion. In *Proceedings of the 18th SIGMORPHON Workshop*, pages 115–125.

Michael Becker and Maria Gouskova. 2016. The typology of sublexical phonotactics. *Phonology*, 33(3):431–465.

Caleb Belth. 2023. Towards a learning-based account of underlying forms: A case study in Turkish. In *Proceedings of the Society for Computation in Linguistics 2023*, pages 332–342, Amherst, MA. Association for Computational Linguistics.

Steven Bird and T. Mark Ellison. 1994. One-level phonology: Autosegmental representations and rules as finite automata. *Computational Linguistics*, 20(1):55–90.

Ellen Broselow. 1976. *The Phonology of Egyptian Arabic*. Ph.D. thesis, University of Massachusetts Amherst.

Ellen Broselow. 2017. Syllable Structure in the Dialects of Arabic. *The Routledge handbook of Arabic linguistics*, pages 32–47.

Ryan Cotterell, Nanyun Peng, and Jason Eisner. 2015. Modeling word forms using latent underlying morphs and phonology. *Transactions of the Association for Computational Linguistics*, 3:433–447.

Nizar Habash and Owen Rambow. 2006. MAGEAD: A morphological analyzer and generator for the Arabic dialects. In *Proceedings of the International Conference on Computational Linguistics and the Conference of the Association for Computational Linguistics (COLING-ACL)*, pages 681–688, Sydney, Australia.

Zellig S. Harris. 1942. Morpheme alternants in linguistic analysis. *Language*, 18(3):169–180.

Mans Hulden, Michael Ginn, Miikka Silfverberg, and Michael Hammond. 2024. PyFoma: a python finite-state compiler module. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 258–265, Bangkok, Thailand. Association for Computational Linguistics.

Hyun Jin Hwangbo. 2018. *The mental representations of allomorphs: an investigation with artificial grammar learning*. Ph.D. thesis, University of Delaware.

Hyun Jin Hwangbo. 2023. Learning allomorphs and their mental representations: An investigation with artificial grammar learning. *Journal of Cognitive Science*, 24(1).

M.J. Kenstowicz and C.W. Kisseberth. 1979. *Generative Phonology: Description and Theory*. Academic Press.

Salam Khalifa, Sarah Payne, Jordan Kodner, Ellen Broselow, and Owen Rambow. 2023. A cautious generalization goes a long way: Learning morphophonological rules. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1793–1805, Toronto, Canada. Association for Computational Linguistics.

Salam Khalifa, Abdelrahim Qaddoumi, Jordan Kodner, and Owen Rambow. 2025. Learning cross-dialectal morphophonology with syllable structure constraints. In *Proceedings of the 12th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 157–167, Abu Dhabi, UAE. Association for Computational Linguistics.

Arya D. McCarthy, Jackson L. Lee, Alexandra DeLucia, Travis Bartley, Milind Agarwal, Lucas F. E. Ashby, Luca Del Signore, Cameron Gibson, Reuben Raff, and Winston Wu. 2023. The sigmorphon 2022 shared task on cross-lingual and low-resource grapheme-to-phoneme conversion. In *Proceedings of the 20th SIGMORPHON Workshop*, pages 230–238.

Joe Pater, Robert Staubs, Karen Jesney, and Brian Smith. 2012. Learning probabilities over underlying representations. In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, pages 62–71, Montréal, Canada. Association for Computational Linguistics.

Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1901–1907, Online. Association for Computational Linguistics.

# A Appendix

| Train size | HMORPH | MTAG-HMORPH | Oracle | NEURAL |
|---|---|---|---|---|
| 100 | 52.6 | 43.9 | 52.6 | 4.1 |
| 200 | 60.6 | 51.1 | 62.4 | 15.3 |
| 300 | 64.5 | 53.6 | 68.0 | 27.1 |
| 400 | 66.9 | 56.8 | 70.3 | 36.2 |
| 500 | 69.0 | 59.5 | 73.1 | 41.9 |
| 600 | 70.7 | 61.6 | 74.2 | 50.7 |
| 700 | 71.4 | 62.9 | 75.6 | 55.2 |
| 800 | 72.7 | 64.9 | 77.5 | 59.5 |
| 900 | 74.1 | 64.9 | 78.9 | 60.0 |
| 1,000 | 75.0 | 66.6 | 79.3 | 66.1 |
| 2,000 | 80.1 | 71.7 | 85.1 | 82.8 |
| **3,000** | **82.9** | 74.5 | 87.1 | 87.7 |
| 4,000 | 84.5 | 76.1 | 89.1 | 89.5 |
| 5,000 | 84.7 | 76.7 | 90.3 | 90.4 |
| 6,000 | 85.0 | 77.6 | 91.3 | 91.1 |
| 7,000 | 84.8 | 78.4 | 91.5 | 91.2 |
| 8,000 | 84.9 | 78.6 | 91.9 | 91.8 |
| 9,000 | 84.9 | 79.0 | 92.1 | 92.2 |
| 10,000 | 84.9 | 79.5 | 92.3 | 92.3 |
| 11,000 | 84.9 | 79.6 | 92.6 | 92.6 |
| 12,000 | 85.0 | 79.7 | 92.9 | 93.0 |
| TRAIN | 85.2 | 79.5 | 93.0 | 92.9 |

Table 3: Learning curve detailed results.
The Oracle presented here is of the HMORPH system.